

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 02-14-02
Searcher: Beverly E. 4994
Terminal time: 40
Elapsed time: _____
CPU time: _____
Total time: 57
Number of Searches: _____
Number of Databases: 3

Search Site

_____ STIC
_____ CM-1
_____ Pre-S

Type of Search

_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors

_____ IG
_____ ☒ STN
_____ ☒ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
_____ ☒ Other CGN

THIS PAGE BLANK (USPTO)

Devi, S.
09/360685

09/360685

FILE 'CAPLUS' ENTERED AT 10:21:36 ON 14 FEB 2002)

L1 6299 SEA FILE=CAPLUS ABB=ON PLU=ON PYLORI
L3 4 SEA FILE=CAPLUS ABB=ON PLU=ON L1 AND (CAI OR (CYTOTOXIN
OR CYTO TOXIN) (W) (ASSOC? OR ASS#) (W) (IMMUNODOMIN? OR
IMMUNO DOMIN?))

-key terms

L3 ANSWER 1 OF 4 CAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 2000:687957 CAPLUS

DOCUMENT NUMBER: 133:247257

TITLE: Method and antisense conjugate for treating
Helicobacter pylori infection

INVENTOR(S): Iversen, Patrick L.; Brand, Randall; Weller,
Dwight D.; Summerton, James E.

PATENT ASSIGNEE(S): AVI BioPharma, Inc., USA; The Board of Regents
of the University of Nebraska

SOURCE: U.S., 21 pp., Cont.-in-part of U.S. 6,030,941.
CODEN: USXXAM

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 3

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 6124271	A	20000926	US 1998-12198	19980123
US 6030941	A	20000229	US 1997-848844	19970430
PRIORITY APPLN. INFO.:			US 1997-36366	P 19970124
			US 1997-848844	A2 19970430
			US 1996-16347	P 19960501
			US 1996-28609	P 19961023

AB A method and conjugate for treating H. pylori infection in
a subject are disclosed. The conjugate is composed of (a) a
nuclease-resistant antisense oligomer effective to inhibit H.
pylori infection in the subject by base-specific
Watson-Crick binding to an H. pylori mRNA transcript, and
(b) a transport moiety conjugated to the oligomer. The transport
moiety is effective to facilitate uptake of the conjugate from the
environment of the stomach into the cytoplasm of H. pylori
cells by active transport or by pH-gradient transport across of the
cell membrane of H. pylori. The conjugate is administered
by oral route, preferably in a swellable polymer bolus designed to
release the conjugate in sustained release.

REFERENCE COUNT: 8 THERE ARE 8 CITED REFERENCES AVAILABLE FOR
THIS RECORD. ALL CITATIONS AVAILABLE IN
THE RE FORMAT

L3 ANSWER 2 OF 4 CAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 1998:527235 CAPLUS

DOCUMENT NUMBER: 129:170512

TITLE: Antisense oligonucleotides conjugates taken up
Helicobacter pylori for treatment of
infection

INVENTOR(S): Iverson, Patrick L.; Bragg, Randall; Weller,
Dwight D.; Summerton, James E.

PATENT ASSIGNEE(S): Antivirals, Inc., USA; University of Nebraska
Board of Regents

SOURCE: PCT Int. Appl., 36 pp.
CODEN: PIXXD2

Searcher : Shears 308-4994

THIS PAGE BLANK (USPTO)

09/360685

DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 3
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9832467	A2	19980730	WO 1998-US1393	19980123
W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG				
US 6030941	A	20000229	US 1997-848844	19970430
AU 9859305	A1	19980818	AU 1998-59305	19980123
EP 973886	A2	20000126	EP 1998-902712	19980123
R: AT, BE, CH, DE, ES, FR, GB, IT, LI, LU, NL				
JP 2001509167	T2	20010710	JP 1998-532174	19980123
PRIORITY APPLN. INFO.:				
			US 1997-36366	P 19970124
			US 1997-848844	A 19970430
			US 1996-16347	P 19960501
			US 1996-28609	P 19961023
			WO 1998-US1393	W 19980123

AB Antisense oligonucleotides inhibiting the expression of *Helicobacter pylori* genes that are conjugated with transport moieties that improve the efficiency of uptake by *H. pylori* are described for use in treatment of stomach infection. The transport moiety facilitates uptake from the environment of the stomach into the cytoplasm of *H. pylori* cells by active transport or by pH-gradient transport across of the cell membrane. The preferred transport moiety is a sugar such as D-galactose or L-arabinose that is taken by *Helicobacter* by active transport but that are not absorbed from the gut.

L3 ANSWER 3 OF 4 CAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 1996:747006 CAPLUS

DOCUMENT NUMBER: 126:55630

TITLE: cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors

AUTHOR(S): Censini, Stefano; Lange, Christina; Xiang, Zhaoying; Crabtree, Jean E.; Ghiara, Paolo; Borodovsky, Mark; Rappuoli, Rino; Covacci, Antonello

CORPORATE SOURCE: Immunobiol. Res. Inst. Siena, Chiron Vaccines, Siena, 53100, Italy

SOURCE: Proc. Natl. Acad. Sci. U. S. A. (1996), 93(25), 14648-14653

CODEN: PNASA6; ISSN: 0027-8424

PUBLISHER: National Academy of Sciences

DOCUMENT TYPE: Journal

LANGUAGE: English

AB CagA, a gene that codes for an immunodominant antigen, is present only in *Helicobacter pylori* strains that are assocd. with

Handwritten note:
 CagA gene...
 US 93-34, 1996 (Baker) Aug.

THIS PAGE BLANK (USPTO)

09/360685

severe forms of gastroduodenal disease (type I strains). We found that the genetic locus that contains *cagA* (*cag*) is part of a 40-kb DNA insertion that likely was acquired horizontally and integrated into the chromosomal glutamate racemase gene. This pathogenicity island is flanked by direct repeats of 31 bp. In some strains, *cag* is split into a right segment (*cagI*) and a left segment (*cagII*) by a novel insertion sequence (IS605). In a minority of *H. pylori* strains, *cagI* and *cagII* are sepd. by an intervening chromosomal sequence. Nucleotide sequencing of the 23,508 base pairs that form the *cagI* region and the extreme 3' end of the *cagII* region reveals the presence of 19 ORFs that code for proteins predicted to be mostly membrane assocd. with one gene (*cagE*), which is similar to the toxin-secretion gene of *Bordetella pertussis*, *ptlC*, and the transport systems required for plasmid transfer, including the *virB4* gene of *Agrobacterium tumefaciens*. Transposon inactivation of several of the *cagI* genes abolishes induction of IL-8 expression in gastric epithelial cell lines. Thus, we believe the *cag* region may encode a novel *H. pylori* secretion system for the export of virulence determinants.

L3 ANSWER 4 OF 4 CAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 1994:2256 CAPLUS

DOCUMENT NUMBER: 120:2256

TITLE: *Helicobacter pylori* proteins useful for vaccines and diagnostics

INVENTOR(S): Covacci, Antonello; Bugnoli, Massimo; Telford, John; Macchia, Giovanni; Rappuoli, Rino

PATENT ASSIGNEE(S): Biocine Sclavo S.p.A., Italy

SOURCE: PCT Int. Appl., 85 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 3

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9318150	A1	19930916	WO 1993-EP472	19930302
W:	AT, AU, BB, BG, BR, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US			
RW:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG			
AU 9336300	A1	19931005	AU 1993-36300	19930302
EP 643770	A1	19950322	EP 1993-905285	19930302
R:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE			
JP 07504565	T2	19950525	JP 1993-515309	19930302
EP 967279	A1	19991229	EP 1999-202698	19930302
R:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE			
JP 2000333686	A2	20001205	JP 2000-126696	19930302
JP 2000350591	A2	20001219	JP 2000-126695	19930302
US 6077706	A	20000620	US 1995-470260	19950606
US 6130059	A	20001010	US 1995-466662	19950606
PRIORITY APPLN. INFO.:			IT 1992-FI52	A 19920302
			WO 1993-EP158	A 19930125
			EP 1993-905285	A3 19930302

Searcher : Shears 308-4994

10/20/2011
10/20/2011
10/20/2011

10/20/2011
10/20/2011

THIS PAGE BLANK (USPTO)

09/360685

JP 1993-515309 A 19930302
WO 1993-EP472 A 19930302
US 1994-256848 B3 19941021

AB The H. **pylori** genes for cytotoxin, **CAI** (**cytotoxin-assocd. immunodominant**) antigen, and heat-shock protein (of the hsp60 family) are cloned and sequenced. The nucleic acids and proteins may be used for diagnosis and the proteins for vaccination against H. **pylori**. The cytotoxin, or CT antigen, gene encoded a 140 kDa protein precursor of a 100 kDa protein with urease-independent vacuolating activity. The **CAI** antigen gene was absent in noncytotoxic H. **pylori** strains. The heterogeneity of the **CAI** antigen appeared to be due to internal duplications in the gene: strain G39 was found to have two identical repeats of a 102 bp sequence within the gene. The heat-shock protein gene was expressed in all H. **pylori** strains tested, both cytotoxic and noncytotoxic. Most sera from patients infected with H. **pylori** and exhibiting gastritis and ulcers contained antibodies to the hsp, but the degree of recognition varied greatly among the patients and the antibody levels did not show any obvious correlation with the type of disease.

(FILE 'MEDLINE, BIOSIS, EMBASE, WPIDS, JICST-EPLUS, JAPIO, TOXLIT, TOXCENTER, PHIC, PHIN' ENTERED AT 10:25:05 ON 14 FEB 2002)

L4
L5

8 S L3

1 DUP REM L4 (1 DUPLICATE REMOVED)

L5 ANSWER 1 OF 7 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
ACCESSION NUMBER: 2001:239751 BIOSIS
DOCUMENT NUMBER: PREV200100239751
TITLE: Helicobacter **pylori** proteins useful for vaccines and diagnostics.
AUTHOR(S): Covacci, Antonello (1); Bugnoli, Massimo; Telford, John; Macchia, Giovanni; Rappuoli, Rino
CORPORATE SOURCE: (1) Vc.Provenzano, 8, 53100, Siena Italy
PATENT INFORMATION: US 6130059 October 10, 2000
SOURCE: Official Gazette of the United States Patent and Trademark Office Patents, (Oct. 10, 2000) Vol. 1239, No. 2, pp. No Pagination. e-file.
ISSN: 0098-1133.
DOCUMENT TYPE: Patent
LANGUAGE: English

AB Helicobacter **pylori** is known to cause or be a cofactor in type B gastritis, peptic ulcers, and gastric tumors. In both developed and developing countries, a high percentage of people are infected with this bacterium. The present invention relates generally to certain H. **pylori** proteins, to the genes which express these proteins, and to the use of these proteins for diagnostic and vaccine applications. Specifically, molecular cloning, nucleotide, and amino acid sequences for the H. **pylori** cytotoxin (CT), the "**Cytotoxin Associated Immunodominant**" (**CAI**) antigen, and the heat shock protein (hsp60). are described herein.

L5 ANSWER 2 OF 7 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
ACCESSION NUMBER: 2001:181462 BIOSIS
DOCUMENT NUMBER: PREV200100181462
TITLE: Helicobacter **pylori** proteins useful for

THIS PAGE BLANK (USPTO)

09/360685

 vaccines and diagnostics.
AUTHOR(S): Covacci, Antonello (1); Bugnoli, Massimo; Telford,
 John; Macchia, Giovanni; Rappuoli, Rino
CORPORATE SOURCE: (1) Siena Italy
 ASSIGNEE: Chiron S.p.A., Siena, Italy
PATENT INFORMATION: US 6090611 July 18, 2000
SOURCE: Official Gazette of the United States Patent and
 Trademark Office Patents, (July 18, 2000) Vol. 1236,
 No. 3, pp. No Pagination. e-file.
 ISSN: 0098-1133.
DOCUMENT TYPE: Patent
LANGUAGE: English

AB A **cytotoxin associated immunodominant**
antigen and the nucleic acid encoding the antigen from *Helicobacter*
pylori are described. This antigen was identified from the
cytotoxin positive CCUG 17874 *Helicobacter pylori* strain,
and both the antigen and the DNA encoding it have been sequenced.
The antigen is a hydrophilic, surface-exposed protein having a
molecular weight of 120-132 kDa. The nucleic acid encoding the
antigen may be incorporated into a vector for transformation of host
cells for expression of the antigen. Both the DNA and the antigen
can be used in assays for detection of disease or infection by
Helicobacter pylori, and may find use in treating and
preventing infection by *Helicobacter pylori* and the
diseases associated with such infection.

L5 ANSWER 3 OF 7 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
ACCESSION NUMBER: 2001:110167 BIOSIS
DOCUMENT NUMBER: PREV200100110167
TITLE: *Helicobacter pylori* proteins useful for
 vaccines and diagnostics.
AUTHOR(S): Covacci, Antonello (1); Bugnoli, Massimo; Telford,
 John; Macchia, Giovanni; Rappuoli, Rino
CORPORATE SOURCE: (1) Siena Italy
 ASSIGNEE: Chiron Corporation
PATENT INFORMATION: US 6077706 June 20, 2000
SOURCE: Official Gazette of the United States Patent and
 Trademark Office Patents, (June 20, 2000) Vol. 1235,
 No. 3, pp. No Pagination. e-file.
 ISSN: 0098-1133.
DOCUMENT TYPE: Patent
LANGUAGE: English

AB *Helicobacter pylori* known to cause or be a cofactor in
type B gastritis, peptic ulcers, and gastric tumors. In both
developed and developing countries, a high percentage of people are
infected with this bacterium. The present invention relates
generally to certain H. **pylori** proteins, to the genes
which express these proteins, and to the use of these proteins for
diagnostic and vaccine applications. Specifically, molecular
cloning, nucleotide, and amino acid sequences for the H.
pylori cytotoxin (CT), the "**Cytotoxin**
Associated Immunodominant" (CAI)
antigen, and the heat shock protein (hsp60) are described herein.

L5 ANSWER 4 OF 7 WPIDS COPYRIGHT 2002 DERWENT INFORMATION LTD
ACCESSION NUMBER: 2001-049948 [06] WPIDS
DOC. NO. NON-CPI: N2001-038270
DOC. NO. CPI: C2001-013762

THIS PAGE BLANK (USPTO)

09/360685

TITLE: Preparing an agent for diagnosis or control of microbial infection, useful particularly against Helicobacter, based on identification of essential genes in defective mutants.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): APFEL, H; FUCHS, T M; GIBBS, C P; HUECK, C J; MEYER, T F
PATENT ASSIGNEE(S): (CREA-N) CREATOGEN GMBH; (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN
COUNTRY COUNT: 93
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG

WO 2000073502	A2	20001207	(200106)*	GE	376
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC					
MW MZ NL OA PT SD SE SL SZ TZ UG ZW					
W: AE AG AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK					
DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP					
KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL					
PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU					
ZA ZW					
DE 19924965	A1	20001207	(200106)		
DE 19927740	A1	20001221	(200106)		
DE 19934029	A1	20010125	(200107)		
AU 2000053998	A	20001218	(200118)		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE

WO 2000073502	A2	WO 2000-EP5024	20000531
DE 19924965	A1	DE 1999-19924965	19990531
DE 19927740	A1	DE 1999-19927740	19990617
DE 19934029	A1	DE 1999-19934029	19990721
AU 2000053998	A	AU 2000-53998	20000531

FILING DETAILS:

PATENT NO	KIND	PATENT NO

AU 2000053998	A Based on	WO 200073502

PRIORITY APPLN. INFO: DE 1999-19934029 19990721; DE 1999-19924965 19990531; DE 1999-19927740 19990617

AN 2001-049948 [06] WPIDS

AB WO 200073502 A UPAB: 20010126

NOVELTY - Preparation of an agent (A) for detection, prevention and/or treatment of microbial infection by:

(i) identifying essential genes (I) and corresponding polypeptides (II);

(ii) identifying compounds that are directed against (II) and inactivate the microbe;

(iii) testing these for suitability for use; and

(iv) formulating selected (A).

DETAILED DESCRIPTION - Identifying essential genes (I) comprises preparation of gene-deficient microorganisms by conditional antisense inhibition (CAI) and/or subtractive

THIS PAGE BLANK (USPTO)

recombination mutagenesis (SRM), then determining viability and/or survival of the deficient organisms.

INDEPENDENT CLAIMS are also included for the following:

- (a) method for identifying essential microbial genes;
- (b) nucleic acid (Ia) representing essential secretory genes of *Helicobacter*, identified by method (a);
- (c) a gene bank containing at least two (Ia);
- (d) a vector containing at least one (Ia) or its fragment;
- (e) cells containing (Ia) or the vector of (d);
- (f) a mutant bank containing at least two microorganisms transformed with the vectors of (d);
- (g) a polypeptide (IIa) encoded by (Ia) and their immunogenic fragments;
- (h) inhibitors (A') that bind specifically to (IIa), or its fragments, and/or modulates its expression, presentation and/or native function;
- (i) method for preparing (IIa) or its fragments;
- (j) antibodies (Ab), or their fragments, that are specific for (IIa); and
- (k) a pharmaceutical composition containing one of (Ia), the vector of (d), the cells of (e), (IIa), Ab or its fragments, or (A).

ACTIVITY - Antibacterial.

MECHANISM OF ACTION - Passive or active immunization.

USE - (A) (which may be a nucleic acid (Ia), vector or host cell containing (Ia), derived polypeptide (IIa), or fragments, (IIa)-specific antibodies or their fragments or an inhibitor of (IIa)) are particularly used for diagnosis, treatment or prevention of infection by *Helicobacter pylori*. Particularly (Ia) and (IIa) are used in DNA, subunit or live vaccines.

ADVANTAGE - The method identifies essential genes, including those that have homologs in other species, so identified (A) should have a broad spectrum of activity. Many gene-deficient cells can be screened quickly, in an automated process, and the identified genes can be used for screening without purification.

Dwg.0/15

L5 ANSWER 5 OF 7 TOXCENTER COPYRIGHT 2002 ACS

ACCESSION NUMBER: 1996:219941 TOXCENTER

COPYRIGHT: Copyright 2002 ACS

DOCUMENT NUMBER: CA12605055630F

TITLE: cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors

AUTHOR(S): Censini, Stefano; Lange, Christina; Xiang, Zhaoying; Crabtree, Jean E.; Ghiara, Paolo; Borodovsky, Mark; Rappuoli, Rino; Covacci, Antonello

CORPORATE SOURCE: Immunobiol. Res. Inst. Siena, Chiron Vaccines, Siena, 53100, Italy.

SOURCE: Proc. Natl. Acad. Sci. U. S. A., (1996) Vol. 93, No. 25, pp. 14648-14653.

CODEN: PNASA6. ISSN: 0027-8424.

COUNTRY: ITALY

DOCUMENT TYPE: Journal

FILE SEGMENT: CAPLUS

OTHER SOURCE: CAPLUS 1996:747006

LANGUAGE: English

ENTRY DATE: Entered STN: 20011116

Last Updated on STN: 20011116

THIS PAGE BLANK (USPTO)

AN 1996:219941 TOXCENTER

CP Copyright 2002 ACS

AB CagA, a gene that codes for an immunodominant antigen, is present only in *Helicobacter pylori* strains that are assocd. with severe forms of gastroduodenal disease (type I strains). We found that the genetic locus that contains cagA (cag) is part of a 40-kb DNA insertion that likely was acquired horizontally and integrated into the chromosomal glutamate racemase gene. This pathogenicity island is flanked by direct repeats of 31 bp. In some strains, cag is split into a right segment (cagI) and a left segment (cagII) by a novel insertion sequence (IS605). In a minority of *H. pylori* strains, cagI and cagII are sepd. by an intervening chromosomal sequence. Nucleotide sequencing of the 23,508 base pairs that form the cagI region and the extreme 3' end of the cagII region reveals the presence of 19 ORFs that code for proteins predicted to be mostly membrane assocd. with one gene (cagE), which is similar to the toxin-secretion gene of *Bordetella pertussis*, *ptlC*, and the transport systems required for plasmid transfer, including the *virB4* gene of *Agrobacterium tumefaciens*. Transposon inactivation of several of the cagI genes abolishes induction of IL-8 expression in gastric epithelial cell lines. Thus, we believe the cag region may encode a novel *H. pylori* secretion system for the export of virulence determinants.

L5 ANSWER 6 OF 7 WPIDS COPYRIGHT 2002 DERWENT INFORMATION LTD
DUPLICATE 1

ACCESSION NUMBER: 1993-303464 [38] WPIDS

CROSS REFERENCE: 1996-485780 [48]

DOC. NO. CPI: C1993-135210

TITLE: Recombinant *Helicobacter pylori* protein and corresp. gene - is a cytotoxin, antigen or heat shock protein used for treating and preventing type B gastritis, gastric ulcers and gastric tumours.

DERWENT CLASS: B04 D16

INVENTOR(S): BUGNOLI, M; COVACCI, A; MACCHIA, G; RAPPUOLI, R;
TELFORD, J

PATENT ASSIGNEE(S): (ISTS) BIOCINE SPA; (CHIR) CHIRON SPA; (CHIR-N) CHIRON SPA; (RICE-N) IST RICERCHE IMMUNOBIOLOGICHE SIENA SRL; (CHIR) CHIRON CORP; (BUGN-I) BUGNOLI M; (COVA-I) COVACCI A; (MACC-I) MACCHIA G; (RAPP-I) RAPPUOLI R; (TELF-I) TELFORD J; (ISTS) BIOCINE SCLAVO SPA

COUNTRY COUNT: 43

PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9318150	A1	19930916	(199338)*	EN	83
RW: AT BE CH DE DK ES FR GB GR IE IT LU MC NL OA PT SE					
W: AT AU BB BG BR CA CH CZ DE DK ES FI GB HU JP KP KR LK LU MG					
MN MW NL NO NZ PL PT RO RU SD SE SK UA US					
AU 9336300	A	19931005	(199405)		
EP 643770	A1	19950322	(199516)	EN	
R: AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE					
JP 07504565	W	19950525	(199529)		
IT 1262895	B	19960722	(199709)		
SG 48353	A1	19980417	(199828)		
US 5928865	A	19990727	(199936)		

THIS PAGE BLANK (USPTO)

09/360685

EP 967279 A1 19991229 (200005) EN
 R: AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE
 US 6077706 A 20000620 (200035)
 US 6090611 A 20000718 (200037)
 US 6130059 A 20001010 (200052)
 JP 2000333686 A 20001205 (200102) 42
 JP 2000350591 A 20001219 (200104) 51

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9318150	A1	WO 1993-EP472	19930302
AU 9336300	A	AU 1993-36300	19930302
EP 643770	A1	EP 1993-905285	19930302
		WO 1993-EP472	19930302
JP 07504565	W	JP 1993-515309	19930302
		WO 1993-EP472	19930302
IT 1262895	B	IT 1992-FI52	19920302
SG 48353	A1	SG 1996-9100	19930302
US 5928865	A Div ex	WO 1993-EP158	19930125
	Div ex	WO 1993-EP472	19930302
	Div ex	US 1994-256848	19941021
	CIP of	US 1995-425194	19950420
	CIP of	US 1995-471491	19950606
		US 1995-477451	19950607
EP 967279	A1 Div ex	EP 1993-905285	19930302
		EP 1999-202698	19930302
US 6077706	A Div ex	WO 1993-EP472	19930302
	Div ex	US 1994-256848	19941021
		US 1995-470260	19950606
US 6090611	A Div ex	WO 1993-EP472	19930302
	Div ex	US 1994-256848	19941021
		US 1995-471491	19950606
US 6130059	A Cont of	WO 1993-EP472	19930302
	Div ex	US 1994-256848	19941021
		US 1995-466662	19950606
JP 2000333686	A Div ex	JP 1993-515309	19930302
		JP 2000-126696	19930302
JP 2000350591	A Div ex	JP 1993-515309	19930302
		JP 2000-126695	19930302

FILING DETAILS:

PATENT NO	KIND	PATENT NO
AU 9336300	A Based on	WO 9318150
EP 643770	A1 Based on	WO 9318150
JP 07504565	W Based on	WO 9318150
EP 967279	A1 Div ex	EP 643770

PRIORITY APPLN. INFO: WO 1993-EP158 19930125; IT 1992-FI52
 19920302

AN 1993-303464 [38] WPIDS

CR 1996-485780 [48]

AB WO 9318150 A UPAB: 20010118

A recombinant Helicobacter pylori protein (I), deriv. or fragment is new.

THIS PAGE BLANK (USPTO)

U.S. PATENT AND
TRADEMARK OFFICE
WASHINGTON, D.C. 20540

Also new are: (a) a vaccine contg. (I); (b) an immunodiagnostic assay, and kit, having at least 1 step involving at least 1 binding partner, (I), opt. labelled or coupled to a solid support; (c) a recombinant polynucleotide encoding (I), a deriv. or fragment; (d) a nucleic acid assay, and kit, and amplification process using (c) as a probe or primer; (e) a vector contg. (c); and (f) a host cell transformed with (e).

(I) is a cytotoxin or a precursor, pref. exhibiting no or reduced toxicity; a **cytotoxin-associated immunodominant** antigen; or a heat shock protein (sequences in specification).

USE/ADVANTAGE - (I) is used to treat, prevent and diagnose H. **pylori** infection, which causes type B gastritis, peptic ulcers and gastric tumours
Dwg.0/5

L5 ANSWER 7 OF 7 TOXLIT
ACCESSION NUMBER: 1994:22745 TOXLIT
DOCUMENT NUMBER: CA-120-002256H
TITLE: Helicobacter **pylori** proteins useful for vaccines and diagnostics.
AUTHOR: Covacci A; Bugnoli M; Telford J; Macchia G; Rappuoli R
SOURCE: (1993). PCT Int. Appl. PATENT NO. 93 18150 09/16/93 (Biocine Sclavo S.p.A.).
PUB. COUNTRY: Italy
DOCUMENT TYPE: Patent
FILE SEGMENT: CA
LANGUAGE: English
OTHER SOURCE: CA 120:2256
ENTRY MONTH: 199405

AB The H. **pylori** genes for cytotoxin, CAI (**cytotoxin-assocd. immunodominant**) antigen, and heat-shock protein (of the hsp60 family) are cloned and sequenced. The nucleic acids and proteins may be used for diagnosis and the proteins for vaccination against H. **pylori**. The cytotoxin, or CT antigen, gene encoded a 140 kDa protein precursor of a 100 kDa protein with urease-independent vacuolating activity. The CAI antigen gene was absent in noncytotoxic H. **pylori** strains. The heterogeneity of the CAI antigen appeared to be due to internal duplications in the gene: strain G39 was found to have two identical repeats of a 102 bp sequence within the gene. The heat-shock protein gene was expressed in all H. **pylori** strains tested, both cytotoxic and noncytotoxic. Most sera from patients infected with H. **pylori** and exhibiting gastritis and ulcers contained antibodies to the hsp, but the degree of recognition varied greatly among the patients and the antibody levels did not show any obvious correlation with the type of disease.

(FILE 'CAPLUS, MEDLINE, BIOSIS, EMBASE, WPIDS, JICST-EPLUS, JAPIO, TOXLIT, TOXCENTER, PHIC, PHIN' ENTERED AT 10:28:44 ON 14 FEB 2002)

L6 359 SEA ABB=ON PLU=ON COVACCI A?/AU
L7 219 SEA ABB=ON PLU=ON BUGNOLI M?/AU
L8 1040 SEA ABB=ON PLU=ON TELFORD J?/AU
L9 228 SEA ABB=ON PLU=ON MACCHIA G?/AU
L10 2220 SEA ABB=ON PLU=ON RAPPUOLI R?/AU
L11 11 SEA ABB=ON PLU=ON L6 AND L7 AND L8 AND L9 AND L10

- Author (S)

Page 1 of 1

THIS PAGE BLANK (USPTO)

09/360685

L12 206 SEA ABB=ON PLU=ON L6 AND (L7 OR L8 OR L9 OR L10)
L13 125 SEA ABB=ON PLU=ON L7 AND (L8 OR L9 OR L10)
L14 322 SEA ABB=ON PLU=ON L8 AND (L9 OR L10)
L15 17 SEA ABB=ON PLU=ON L9 AND L10
L16 3396 SEA ABB=ON PLU=ON L6 OR L7 OR L8 OR L9 OR L10
L17 9 SEA ABB=ON PLU=ON (L12 OR L13 OR L14 OR L16) AND (CAI
OR (CYTOTOXIN OR CYTO TOXIN) (W) (ASSOC? OR ASS#) (W) (IMMUNO
DOMIN? OR IMMUNO DOMIN?))
L18 19 SEA ABB=ON PLU=ON L11 OR L15 OR L17
L19 11 DUP REM L18 (8 DUPLICATES REMOVED)

L19 ANSWER 1 OF 11 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
ACCESSION NUMBER: 2001:239751 BIOSIS
DOCUMENT NUMBER: PREV200100239751
TITLE: Helicobacter pylori proteins useful for vaccines and
diagnostics.
AUTHOR(S): Covacci, Antonello (1); Bugnoli,
Massimo; Telford, John; Macchia,
Giovanni; Rappuoli, Rino
CORPORATE SOURCE: (1) Vc.Provenzano, 8, 53100, Siena Italy
PATENT INFORMATION: US 6130059 October 10, 2000
SOURCE: Official Gazette of the United States Patent and
Trademark Office Patents, (Oct. 10, 2000) Vol. 1239,
No. 2, pp. No Pagination. e-file.
ISSN: 0098-1133.

DOCUMENT TYPE: Patent
LANGUAGE: English

AB Helicobacter pylori is known to cause or be a cofactor in type B
gastritis, peptic ulcers, and gastric tumors. In both developed and
developing countries, a high percentage of people are infected with
this bacterium. The present invention relates generally to certain
H. pylori proteins, to the genes which express these proteins, and
to the use of these proteins for diagnostic and vaccine
applications. Specifically, molecular cloning, nucleotide, and amino
acid sequences for the H. pylori cytotoxin (CT), the "
Cytotoxin Associated Immunodominant" (CAI) antigen, and the heat shock protein (hsp60). are
described herein.

L19 ANSWER 2 OF 11 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
ACCESSION NUMBER: 2001:181462 BIOSIS
DOCUMENT NUMBER: PREV200100181462
TITLE: Helicobacter pylori proteins useful for vaccines and
diagnostics.
AUTHOR(S): Covacci, Antonello (1); Bugnoli,
Massimo; Telford, John; Macchia,
Giovanni; Rappuoli, Rino
CORPORATE SOURCE: (1) Siena Italy
ASSIGNEE: Chiron S.p.A., Siena, Italy
PATENT INFORMATION: US 6090611 July 18, 2000
SOURCE: Official Gazette of the United States Patent and
Trademark Office Patents, (July 18, 2000) Vol. 1236,
No. 3, pp. No Pagination. e-file.
ISSN: 0098-1133.

DOCUMENT TYPE: Patent
LANGUAGE: English

AB A cytotoxin associated immunodominant
antigen and the nucleic acid encoding the antigen from Helicobacter

THIS PAGE BLANK (USPTO)

pylori are described. This antigen was identified from the cytotoxin positive CCUG 17874 *Helicobacter pylori* strain, and both the antigen and the DNA encoding it have been sequenced. The antigen is a hydrophilic, surface-exposed protein having a molecular weight of 120-132 kDa. The nucleic acid encoding the antigen may be incorporated into a vector for transformation of host cells for expression of the antigen. Both the DNA and the antigen can be used in assays for detection of disease or infection by *Helicobacter pylori*, and may find use in treating and preventing infection by *Helicobacter pylori* and the diseases associated with such infection.

L19 ANSWER 3 OF 11 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.

ACCESSION NUMBER: 2001:110167 BIOSIS

DOCUMENT NUMBER: PREV200100110167

TITLE: *Helicobacter pylori* proteins useful for vaccines and diagnostics.

AUTHOR(S): Covacci, Antonello (1); Bugnoli, Massimo; Telford, John; Macchia, Giovanni; Rappuoli, Rino

CORPORATE SOURCE: (1) Siena Italy
ASSIGNEE: Chiron Corporation

PATENT INFORMATION: US 6077706 June 20, 2000

SOURCE: Official Gazette of the United States Patent and Trademark Office Patents, (June 20, 2000) Vol. 1235, No. 3, pp. No Pagination. e-file.
ISSN: 0098-1133.

DOCUMENT TYPE: Patent

LANGUAGE: English

AB *Helicobacter pylori* known to cause or be a cofactor in type B gastritis, peptic ulcers, and gastric tumors. In both developed and developing countries, a high percentage of people are infected with this bacterium. The present invention relates generally to certain *H. pylori* proteins, to the genes which express these proteins, and to the use of these proteins for diagnostic and vaccine applications. Specifically, molecular cloning, nucleotide, and amino acid sequences for the *H. pylori* cytotoxin (CT), the "Cytotoxin Associated Immunodominant" (CAI) antigen, and the heat shock protein (hsp60) are described herein.

L19 ANSWER 4 OF 11 CAPLUS COPYRIGHT 2002 ACS DUPLICATE 1

ACCESSION NUMBER: 1996:747006 CAPLUS

DOCUMENT NUMBER: 126:55630

TITLE: cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors

AUTHOR(S): Censini, Stefano; Lange, Christina; Xiang, Zhaoying; Crabtree, Jean E.; Ghiara, Paolo; Borodovsky, Mark; Rappuoli, Rino; Covacci, Antonello

CORPORATE SOURCE: Immunobiol. Res. Inst. Siena, Chiron Vaccines, Siena, 53100, Italy

SOURCE: Proc. Natl. Acad. Sci. U. S. A. (1996), 93(25), 14648-14653

CODEN: PNASA6; ISSN: 0027-8424

PUBLISHER: National Academy of Sciences

DOCUMENT TYPE: Journal

LANGUAGE: English

THIS PAGE BLANK (USPTO)

AB CagA, a gene that codes for an immunodominant antigen, is present only in *Helicobacter pylori* strains that are assocd. with severe forms of gastroduodenal disease (type I strains). We found that the genetic locus that contains cagA (cag) is part of a 40-kb DNA insertion that likely was acquired horizontally and integrated into the chromosomal glutamate racemase gene. This pathogenicity island is flanked by direct repeats of 31 bp. In some strains, cag is split into a right segment (cagI) and a left segment (cagII) by a novel insertion sequence (IS605). In a minority of *H. pylori* strains, cagI and cagII are sepd. by an intervening chromosomal sequence. Nucleotide sequencing of the 23,508 base pairs that form the cagI region and the extreme 3' end of the cagII region reveals the presence of 19 ORFs that code for proteins predicted to be mostly membrane assocd. with one gene (cagE), which is similar to the toxin-secretion gene of *Bordetella pertussis*, ptlC, and the transport systems required for plasmid transfer, including the virB4 gene of *Agrobacterium tumefaciens*. Transposon inactivation of several of the cagI genes abolishes induction of IL-8 expression in gastric epithelial cell lines. Thus, we believe the cag region may encode a novel *H. pylori* secretion system for the export of virulence determinants.

L19 ANSWER 5 OF 11 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
DUPLICATE 2

ACCESSION NUMBER: 1994:325647 BIOSIS
DOCUMENT NUMBER: PREV199497338647
TITLE: Virulence determinants in *Helicobacter pylori* infection.
AUTHOR(S): Covacci, A. (1); Censini, S. (1);
Telford, J. L. (1); Bugnoli, M. (1);
Burrioni, D. (1); Dell'orco, M.; Ghiara, P. (1);
Xiang, Z. (1); Macchia, G. (1); Papini, E.;
Montecucco, C.; Rappuoli, R. (1)
CORPORATE SOURCE: (1) Immunobiol. Res. Inst. Siena, Via Fiorentina 1,
53100 Siena Italy
SOURCE: Freer, J. [Editor]; Aitken, R. [Editor]; Alouf, J. E.
[Editor]; Boulnois, G. [Editor]. FEMS Symposium,
(1994) No. 73, pp. 43-50. FEMS Symposium; Bacterial
protein toxins.
Publisher: Gustav Fischer Verlag Wollgrasweg 49,
D-7000 Stuttgart, Germany.
Meeting Info.: Sixth European Workshop Stirling,
Scotland, UK June 27-July 2, 1993
ISSN: 0163-9188. ISBN: 3-437-11535-9, 1-56081-385-7.
DOCUMENT TYPE: Book; Conference
LANGUAGE: English

L19 ANSWER 6 OF 11 CAPLUS COPYRIGHT 2002 ACS DUPLICATE 3
ACCESSION NUMBER: 1994:2256 CAPLUS
DOCUMENT NUMBER: 120:2256
TITLE: *Helicobacter pylori* proteins useful for vaccines
and diagnostics
INVENTOR(S): Covacci, Antonello; Bugnoli,
Massimo; Telford, John;
Macchia, Giovanni; Rappuoli,
Rino
PATENT ASSIGNEE(S): Biocine Sclavo S.p.A., Italy
SOURCE: PCT Int. Appl., 85 pp.

THIS PAGE BLANK (USPTO)

09/360685

CODEN: PIXXD2
 DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 3
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9318150	A1	19930916	WO 1993-EP472	19930302
W: AT, AU, BB, BG, BR, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US				
RW: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG				
AU 9336300	A1	19931005	AU 1993-36300	19930302
EP 643770	A1	19950322	EP 1993-905285	19930302
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE				
JP 07504565	T2	19950525	JP 1993-515309	19930302
EP 967279	A1	19991229	EP 1999-202698	19930302
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE				
JP 2000333686	A2	20001205	JP 2000-126696	19930302
JP 2000350591	A2	20001219	JP 2000-126695	19930302
US 6077706	A	20000620	US 1995-470260	19950606
US 6130059	A	20001010	US 1995-466662	19950606

PRIORITY APPLN. INFO.:

IT 1992-FI52	A	19920302
WO 1993-EP158	A	19930125
EP 1993-905285	A3	19930302
JP 1993-515309	A	19930302
WO 1993-EP472	A	19930302
US 1994-256848	B3	19941021

AB The H. pylori genes for cytotoxin, **CAI** (cytotoxin-assocd. immunodominant) antigen, and heat-shock protein (of the hsp60 family) are cloned and sequenced. The nucleic acids and proteins may be used for diagnosis and the proteins for vaccination against H. pylori. The cytotoxin, or CT antigen, gene encoded a 140 kDa protein precursor of a 100 kDa protein with urease-independent vacuolating activity. The **CAI** antigen gene was absent in noncytotoxic H. pylori strains. The heterogeneity of the **CAI** antigen appeared to be due to internal duplications in the gene: strain G39 was found to have two identical repeats of a 102 bp sequence within the gene. The heat-shock protein gene was expressed in all H. pylori strains tested, both cytotoxic and noncytotoxic. Most sera from patients infected with H. pylori and exhibiting gastritis and ulcers contained antibodies to the hsp, but the degree of recognition varied greatly among the patients and the antibody levels did not show any obvious correlation with the type of disease.

L19 ANSWER 7 OF 11 TOXLIT
 ACCESSION NUMBER: 1994:22745 TOXLIT
 DOCUMENT NUMBER: CA-120-002256H
 TITLE: Helicobacter pylori proteins useful for vaccines and diagnostics.
 AUTHOR: Covacci A; Bugnoli M;
 Telford J; Macchia G; Rappuoli
 R

THIS PAGE BLANK (USPTO)

09/360685

SOURCE: (1993). PCT Int. Appl. PATENT NO. 93 18150 09/16/93
(Biocine Sclavo S.p.A.).
PUB. COUNTRY: Italy
DOCUMENT TYPE: Patent
FILE SEGMENT: CA
LANGUAGE: English
OTHER SOURCE: CA 120:2256
ENTRY MONTH: 199405

AB The *H. pylori* genes for cytotoxin, **CAI** (cytotoxin -assocd. immunodominant) antigen, and heat-shock protein (of the hsp60 family) are cloned and sequenced. The nucleic acids and proteins may be used for diagnosis and the proteins for vaccination against *H. pylori*. The cytotoxin, or CT antigen, gene encoded a 140 kDa protein precursor of a 100 kDa protein with urease-independent vacuolating activity. The **CAI** antigen gene was absent in noncytotoxic *H. pylori* strains. The heterogeneity of the **CAI** antigen appeared to be due to internal duplications in the gene: strain G39 was found to have two identical repeats of a 102 bp sequence within the gene. The heat-shock protein gene was expressed in all *H. pylori* strains tested, both cytotoxic and noncytotoxic. Most sera from patients infected with *H. pylori* and exhibiting gastritis and ulcers contained antibodies to the hsp, but the degree of recognition varied greatly among the patients and the antibody levels did not show any obvious correlation with the type of disease.

L19 ANSWER 8 OF 11 EMBASE COPYRIGHT 2002 ELSEVIER SCI. B.V.
ACCESSION NUMBER: 93176262 EMBASE
DOCUMENT NUMBER: 1993176262.
TITLE: Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer.
AUTHOR: Covacci A.; Censini S.; Bugnoli M.; Petracca R.; Burrone D.; **Macchia G.**; Massone A.; Papini E.; Xiang Z.; Figura N.; **Rappuoli R.**
CORPORATE SOURCE: Immunobiological Res. Inst. Siena, Via Fiorentina 1, 53100 Siena, Italy
SOURCE: Proceedings of the National Academy of Sciences of the United States of America, (1993) 90/12 (5791-5795).
ISSN: 0027-8424 CODEN: PNASA6
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 004 Microbiology
026 Immunology, Serology and Transplantation
048 Gastroenterology
LANGUAGE: English
SUMMARY LANGUAGE: English

AB *Helicobacter pylori* has been associated with gastritis, peptic ulcer, and gastric adenocarcinoma. We report the nucleotide sequence and expression of an immunodominant antigen of *H. pylori* and the immune response to the antigen during disease. The antigen, named CagA (cytotoxin-associated gene A), is a hydrophilic, surface-exposed protein of 128 kDa produced by most clinical isolates. The size of the cagA gene and its protein varies in different strains by a mechanism that involves duplication of regions within the gene. Clinical isolates that do not produce the antigen do not have the gene and are unable to produce an active

THIS PAGE BLANK (USPTO)

09/360685

vacuolating cytotoxin. An ELISA to detect the immune response against a recombinant fragment of this protein detects 75.3% of patients with gastroduodenal diseases and 100% of patients with duodenal ulcer ($P < 0.0005$), suggesting that only bacteria harboring this protein are associated with disease.

L19 ANSWER 9 OF 11 CAPLUS COPYRIGHT 2002 ACS DUPLICATE 4
ACCESSION NUMBER: 1994:6614 CAPLUS
DOCUMENT NUMBER: 120:6614
TITLE: The Hsp60 protein of Helicobacter pylori:
structure and immune response in patients with
gastroduodenal diseases
AUTHOR(S): **Macchia, Giovanni**; Massone, Annalisa;
Burroni, Daniela; Covacci, Antonello; Censini,
Stefano; **Rappuoli, Rino**
CORPORATE SOURCE: IRIS Immunobiol. Res. Inst. Siena, Siena,
I-53100, Italy
SOURCE: Mol. Microbiol. (1993), 9(3), 645-52
CODEN: MOMIEE; ISSN: 0950-382X
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Helicobacter pylori is a human pathogen that has been assocd. with gastritis, peptic ulcer and gastric carcinoma. The role of the direct action of H. pylori virulence factors and of the induction of autoreactive immunity in the development of chronic gastritis has not been clarified yet. Here the authors report the cloning and mol. characterization of a gene of H. pylori coding for a protein of 58 kDa, recognized by sera of patients affected by H. pylori-induced gastroduodenal diseases. This antigen is present in all the H. pylori strains tested and it belongs to the Hsp60 family of heat-shock proteins, with high homol. with other bacterial and eukaryotic proteins of the same family. This class of homologous proteins has been implicated in the induction of autoimmune disorders in different systems. The presence in infected patients of anti-H. pylori Hsp60 antibodies, potentially cross-reacting with the human homolog, and cross-reactivity between human Hsp60 and a rabbit antiserum against H. pylori Hsp60 suggest that a role of this protein in gastroduodenal diseases is possible.

L19 ANSWER 10 OF 11 JAPIO COPYRIGHT 2002 JPO
ACCESSION NUMBER: 2000-350591 JAPIO
TITLE: VACCINE AND HELICOBACTER PYLORI PROTEIN USEFUL
IN DIAGNOSIS
INVENTOR: **COVACCI ANTONELLO; BUGNOLI
MASSIMO; TELFORD JOHN;
MACCHIA GIOVANNI; RAPPUOLI RINO**
PATENT ASSIGNEE(S): CHIRON SPA)
PATENT INFORMATION:

PATENT NO	KIND	DATE	ERA	MAIN IPC
JP 2000350591A		20001219	Heisei	C12N015-09

JP
APPLICATION INFORMATION
ST19N FORMAT: JP1993-126695 19930302
ORIGINAL: JP2000126695 Heisei
PRIORITY APPLN. INFO.: IT1992FI 52 19920302

Searcher : Shears 308-4994

THIS PAGE BLANK (USPTO)

09/360685

PRIORITY APPLN. INFO.: WO1993EP 9300158 19930125
SOURCE: PATENT ABSTRACTS OF JAPAN (CD-ROM), Unexamined
Applications, Vol. 2000

AN 2000-350591 JAPIO

AB PROBLEM TO BE SOLVED: To obtain a novel Helicobacter pylori protein which possesses a specific sequence containing a specific number of successive amino acids, can be linked with an antibody against Helicobacter pylori and is useful in the diagnosis or prophylaxis of the above bacterial infection and as a vaccine for treatment.

SOLUTION: This Helicobacter pylori protein is a novel recombinant polypeptide which comprises a sequence containing at least 8 successive amino acids and is useful in the diagnosis or prophylaxis of the above bacterial infection and as a vaccine for treatment. The sequence containing at least 8 successive amino acids includes at least a part which can be linked with an antibody against Helicobacter pylori in this sequence consisting of at least 8 successive amino acids or in this polypeptide and is obtained from the amino acid sequence of the formula. This recombinant polypeptide is obtained by preparing the above bacteria-derived DNA library, screening out this library by a partial sequence and expressing the resultant gene.

COPYRIGHT: (C)2000, JPO

L19 ANSWER 11 OF 11 JAPIO COPYRIGHT 2002 JPO

ACCESSION NUMBER: 2000-333686 JAPIO
TITLE: VACCINE AND HELICOBACTER PYLORI PROTEIN USEFUL
FOR DIAGNOSIS

INVENTOR: COVACCI ANTONELLO; BUGNOLI
MASSIMO; TELFORD JOHN;
MACCHIA GIOVANNI; RAPPUOLI RINO

PATENT ASSIGNEE(S): CHIRON SPA)

PATENT INFORMATION:

PATENT NO	KIND	DATE	ERA	MAIN IPC
JP 2000333686A		20001205	Heisei	C12N015-09

JP

APPLICATION INFORMATION

ST19N FORMAT: JP1993-126696 19930302
ORIGINAL: JP2000126696 Heisei
PRIORITY APPLN. INFO.: IT1992FI 52 19920302
PRIORITY APPLN. INFO.: WO1993EP 9300158 19930125
SOURCE: PATENT ABSTRACTS OF JAPAN (CD-ROM), Unexamined
Applications, Vol. 2000

AN 2000-333686 JAPIO

AB PROBLEM TO BE SOLVED: To obtain a new recombinant polypeptide having a specific amino acid sequence, consisting of an antigenic protein of Helicobacter pylori, having immunogenicity and useful for a vaccine for the treatment, the diagnosis and the like of infection of the above bacterium.

SOLUTION: This recombinant polypeptide having immunogenicity, and exhibiting no function or substantially a low function as a toxin has the following characteristics: a new recombinant polypeptide containing a continuous sequence of at least 8 amino acids; the continuous sequence of at least 8 amino acids contains at least one site; the site can be bound to an antibody against Helicobacter pylori both in the sequence consisting of the continuous sequence of

THIS PAGE BLANK (USPTO)

09/360685

at least 8 amino acids and in the polypeptide; the continuous sequence of at least 8 amino acids is obtained from the amino acid sequence of the formula; and this polypeptide is useful for a vaccine, the diagnosis or the like of the infection of the above bacterium.

COPYRIGHT: (C)2000,JPO

=> fil hom

FILE 'HOME' ENTERED AT 10:36:40 ON 14 FEB 2002

THIS PAGE BLANK (USPTO)

SUMMARIES

Result :No.	Query No.	Score	Query		Length	DB	ID	Description
			Match	%				
1	5867	100.0	1147	14	AAR41199		CAI antigen. Heli	
2	5867	100.0	1338	19	AAR80600		Helicobacter pylor	
3	5281	90.0	1181	15	AAR53269		Tag A antigen of H	
4	5281	90.0	1181	16	AAR72593		H. pylori taga ant	
5	5281	90.0	1181	17	AAR93107		Helicobacter pylor	
6	5086	86.7	1178	18	AAR20174		H. pylori cytoplas	
7	5086	86.7	1183	18	AAR52084		H. pylori cytoplas	
8	4056.5	69.1	859	16	AAR53268		120-128 kilodalton	
9	4056.5	69.1	859	16	AAR72594		H. pylori taga ant	
10	4031.5	68.7	850	17	AAR91308		Helicobacter pylor	
11	1591	27.1	318	20	AAR99997		Expressed antigen	

12	917	15.6	187	20	AAW89878
13	326.5	5.6	5024	22	AAW82935
14	307	5.2	1254	11	AAW07503
15	307	5.2	1254	18	AAW24575
16	294	5.0	2688	22	AAW40883
17	291	5.0	2663	22	AAW39097
18	290	4.9	1979	21	AAW18171
19	284	4.8	2285	20	AAW98149
20	281.5	4.8	1279	22	AAW83047
21	275.5	4.7	2485	21	AAW18172
22	257	4.4	2954	20	AAW01632
23	255	4.3	941	19	AAW68205
24	255	4.3	1392	20	AAW06999
25	253	4.3	1411	17	AAW02258
26	252.5	4.3	980	21	AAW18294
27	249.5	4.3	2295	21	AAW18180
28	249	4.2	1639	19	AAW54145
29	248.5	4.2	1312	18	AAW22775
30	248.5	4.2	1312	19	AAW71295
31	248	4.2	1095	22	AAW83030
32	247.5	4.2	1972	17	AAW00024
33	246.5	4.2	1822	13	AAW27745
34	246	4.2	1119	20	AAV19934
35	246	4.2	1354	18	AAW23654
36	246	4.2	1354	19	AAW71020
37	246	4.2	1354	20	AAW07082
38	245.5	4.2	975	22	AAW94042
39	243.5	4.2	1087	20	AAV19935
40	242	4.1	665	21	AAW18278
41	242	4.1	1427	12	AAW10534
42	241.5	4.1	1654	6	AAW50777
43	241.5	4.1	1788	22	AAW40467
44	241.5	4.1	1886	19	AAW54241
45	240.5	4.1	1714	21	AAW18275

ALIGNMENTS

RESULT	1
AAR41199	
IID	AAR41199 standard; Protein; I147 AA.
XX	
AC	AAR41199;
XX	
DT	17-MAR-1994 (first entry)
XX	
DE	CAI antigen.
XX	
KW	Cytotoxin; CT; H. pylori; precursor; vacuol
KW	eukaryote; cytotoxin-associated immunodomin
KW	heat shock protein; hsp; hsp60; type B; gas
KW	gastric tumours.
XX	
OS	Helicobacter pylori.
XX	
FH	Location/Qualifiers
FT	600..900
FT	/note= "Hydrophilic region"
FT	703..714
FT	/note= "Repeat region"
FT	748..759
FT	/note= "Repeat region"
FT	880..885
FT	/note= "Asparagine rich reg
FT	890..894
FT	/note= "Repeat region"
FT	909..913
FT	/note= "Repeat region"
FT	942..956
FT	/note= "Peptide D3"
XX	
PN	W09318150-A.

XX	PD	16-SEP-1993.	QY	301	SVLMGSHNGIEPEKVSLLYGGNGPGARHDWATVGYKDQGNVATIIINVMKNGSLV	360	
XX	PF	02-MAR-1993; 93WO-EP00472.	Db	301	svlmgshngiepeksvlllyggngpgarhdwnatvgykdqgnvatiiinvhmknsgslv	360	
XX	PR	02-MAR-1993; 92IT-QF10052.	QY	361	IAGGEGKINNPSPFYLLKEDQLTGSORALSQBEIQNKIDPMEFLAQNNAKLDNLSEKEK	420	
XX	PR	25-JAN-1993; 93WO-EP00158.	Db	361	iagggegknnpssfylykedqltgsqralsqeeelqnlkldfmeqlagnakldnlsekekek	420	
XX	PA	(BIOC-) BIOCINE SCLAVO SPA.	QY	421	FTELEKDFOKDSKAYLDALGNDRIAFVSKKDTKHSALITEFGNGDLISYTLKDYGGKADKA	480	
XX	PI	Bugnoli M, Covacci A, Macchia G, Rappuoli R, Telford J;	Db	421	ftelekdfokdsakayldalgnndriaafvskkdkthsalitefngdlsytlkygkkaadka	480	
XX	DR	WPI: 1993-303464/38.	QY	481	LDREKNVTLQGLKHDGVNFVDSYFNKYTNASKNPKGVGTNGVSHLEVGFNKVAIFNL	540	
XX	DR	N-PSDB; AAQ48733.	Db	481	ldreknvltqgskhdgvnmfvdysfnkytnasknkpknkgvgtngvshlevgfknvaifnl	540	
XX	PT	Recombinant Helicobacter pylori protein and corresp. gene - is a	QY	541	PDNLNLAITSFVRRNLEDKLTTKGLSPQANKLIKDFLSSNKELVGKTLNFKAVADAKN	600	
XX	PT	cytotoxin, antigen or heat shock protein used for treating and	Db	541	pdnlmlaitsfvrnledklttkglspqeanlkldfissnkelygktnlfnkavadakn	600	
XX	PT	preventing type B gastritis, gastric ulcers and gastric tumours	QY	601	TGNYDEVKKAQDKLESLRKRHELEKVEKKLESKSGNKNKMEAKAQAQNSOKDEIFALIN	660	
XX	PS	Claim 5; Fig 4; 83pp; English.	Db	601	tgnydevkkaqdkleslrkrhelekevekklesksgnknkmeakaqaqnsqkdeifalin	660	
XX	CC	This sequence represents the cytotoxin-associated immunodominant	QY	661	KEANRDARAIAQAQNLKGIKRELSKLENVKNLKDFOKSFDEFKNGKNKDFSKAEETLK	720	
XX	CC	antigen, CAI, of H. pylori. The protein is a hydrophilic,	Db	661	keanrdaraiayaqnlkgikrelsklenvknldfoksfdefkngknkdfskaeetlk	720	
XX	CC	surface-exposed protein having a molecular weight of approximately	QY	721	ALKGVSKDLGINPEWISKVENLNAALNFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV	780	
XX	CC	120-132 kD, pI 128-130 kD, and an isoelectric point of 9.72. H.	Db	721	alkgvskdldginpewiskvenlnaalnfnkngknkdfskvtqakadlensvkdviinqkv	780	
XX	CC	pylori isolates which do not produce the CAI antigen do not have the	QY	781	TDKVDNLNAQVSAKATGDSRVEQALADLNKFSKEQLAQQAQKNESLNARKKSEIYQSV	840	
XX	CC	cai gene, and are also unable to produce an active cytotoxin. The	Db	781	tdkvdnlnaqvsvakatgdsrveqaladlnkfskeqlaqqaqkneslnarkkseliyqsv	840	
XX	CC	association between the presence of the cai gene and cytotoxicity	QY	841	KNGVNGTUVGLNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNGLKNEPIYAKVNKKK	900	
XX	CC	suggests that the product of the cai gene is necessary for the	Db	841	kngvngtluvgnglsqaeattlsknfstdikkelnaklgnfnnnnnglknepiyakvnkkk	900	
XX	CC	transcription, folding, export or function of the cytotoxin.	QY	901	AGQAASLEETPYAQVAKVKNNAKIDRLNQTASGLGVGVAAGFPPLKRHRDKVDLSDKVLGR	960	
XX	CC	Alternatively both the cai gene and the ct gene are absent in	Db	901	agqaasleepiyavakvknakidrlnqlasglgvvgaagfpplkrhrdkvdlsvglgr	960	
XX	CC	noncytotoxic strains suggesting physical linkage between the genes.	QY	961	NOELAQKIDNLNQAQVSEAKAGFFGNLEQTIDKLKSTTHNPNMLWVESAKKVPASLSAKL	1020	
XX	CC	The absence of a typical leader sequence suggests the presence of an	Db	961	nqelaqkldnlngqvseakagffgnleqtdikldksthmpmnlwvesakvpsalsakl	1020	
XX	CC	independent export system. Computer searches for promoter regions in	QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC	the region upstream from the ATG codon identified sequences resembling	Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC	either -10 or -35 regions, however a region with a good consensus to	QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC	the E. coli or H. pylori promoter sequences was not found. The CAI	Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC	antigen is very hydrophilic and does not show obvious leader peptide	QY	1141	KGGFQKS 1147		
XX	CC	or transmembrane sequences. The most hydrophilic region is from amino	Db	1141	kggfqs 1147		
XX	CC	acids 600-900, where a number of unusual features are observed. There	QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC	is a repetition of the sequences EFKNGKNKDFSK and EPVIA, and the	Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC	presence of a stretch of six contiguous asparagines. This protein,	QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC	and others derived from H. pylori, esp. cytotoxin (CT) or a heat shock	Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC	protein (see also AAR41198 and AAR41200), may be used to treat, prevent	QY	1141	KGGFQKS 1147		
XX	CC	and diagnose H. pylori infection. H. pylori is the causative agent of	Db	1141	kggfqs 1147		
XX	CC	type B gastritis, peptic ulcers and gastric tumours.	QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC	Sequence 1147 AA;	Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs 1147		
XX	CC		QY	1021	DNYATNSHIRINSNIKNGAINEKATGMLTKNPEWKLVDKIVAHNVGSPVLSYDKIG	1080	
XX	CC		Db	1021	dnyatnshirinsnikngaineakatgmltknpewklvndkivahnvgsvpvlsydkig	1080	
XX	CC		QY	1081	FNQKNMKDYSDFSTFKLNNAVADTNSGFTQFLTNAFSTASYCYCLARENAEHGKKNVT	1140	
XX	CC		Db	1081	fnqknmkdysdfstfklnnavadtngsgftqfltnafstasyyclarenaehgknvnt	1140	
XX	CC		QY	1141	KGGFQKS 1147		
XX	CC		Db	1141	kggfqs		

KW ulcer; stomach cancer.
XX Chimeric - Helicobacter pylori.
OS Chimeric - Vibrio cholerae.
OS Chimeric - Vibrio cholerae.
FH Key Location/Qualifiers
FT Protein 1..1147
FT /label= CagA
FT 1148..1338
FT /label= A2+B_subunits
XX W09844130-A1.
XX 08-OCT-1998.
XX 31-MAR-1998; 98WO-KR000073.
XX 31-MAR-1997; 97KR-0011951.
XX 31-MAR-1997; 97KR-0011950.
XX (DAEW-) DAEWOONG PHARM CO LTD.
XX Choi D, Jung H, Kim B, Park M, Shin S, Yu Y;
XX WPI; 1998-568279/48.
XX N-PSDB; AAV62461.
XX New chimeric proteins for use against Helicobacter pylori
PT comprising an antigenic protein of H. pylori and A1 and B subunits
PT of Vibrio cholerae toxin, preferably produced by recombinant
PT techniques
XX Claim 8; Page 83; 102pp; English.
XX This polypeptide comprises a protein fusion between the CagA
CC antigen of Helicobacter pylori and the A2 and B subunits of the
CC Vibrio cholerae toxin. A fusion gene (see AAV62461) encoding the
CC polypeptide is also claimed. The invention relates to: chimeric
CC proteins comprising antigenic proteins of H. pylori and A2 and B
CC subunits of V. cholerae toxin; recombinant DNAs encoding such
CC chimeric proteins; recombinant expression vectors; a process for
CC preparing the chimeric proteins using transformed microbial host
CC cells; and to preventative and therapeutic vaccines comprising the
CC chimeric proteins for H. pylori-associated diseases such as
CC gastritis, gastric ulcer, duodenal ulcer and gastric cancer
CC (claimed). The chimeric proteins are designed to have excellent
CC immunogenicity, to be stable in the stomach, to penetrate the
CC mucous membrane of the intestine, and to stimulate production of
CC sIgA. They can additionally be used as active ingredients in
CC diagnostic kits for H. pylori infection, and for production of
XX anti-H. pylori antibody.
XX Sequence 1338 AA;
Query Match 100.0%; Score 5867; DB 19; Length 1338;
Best Local Similarity 100.0%; Pred. No. 1.3e-294;
Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTNETIDQOQTEAFAFNQOQFINNLOVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
DB 1 mtnetiddgqpteaafnpqgfinnliqvaflkvdnavasyddpdkpivdkndrnrqafeg 60
QY 61 ISQLREESYNAIKATPKKNQYFDFINKSNDLINKNDLIDVESSTKSFQKFGDQYRIF 120
DB 61 isqlreesylnkaikptknqyfsdfinksndlinkndlidvesstksfqkfgdqyryrif 120
QY 121 TSWVSHQNDPSKINTSRINFRNMENTIQPPILDKKEAFLKSAKQSFAGIIIGNQIRTDQ 180
DB 121 tswvshqndpskintsrinfrnmentiqppilddkkaeflksakqsfagiignqirtdd 180
QY 181 KPMGVFDESLSKERQAEKNGEPTGGDWLIDIFLSFTFDKKSQSDVKEAINQEPVPHVQDDI 240
DB 181 kpmgvfdeslskerqaeakngeptggdwldiflsftfdkksqsdvkeainqepvphvqpddi 240

DB 181 kfmgvfdeslskerqaeakngeptggdwldiflsftfdkksqsdvkeainqepvphvqpddi 240
QY 241 ATTTTDIOGLPEARDLDERGNFSKFTLGDWEMLDVEGVADIDPNYKFNOLLHNHNS 300
DB 241 atttdioglpeardlldergnfskftlgdmemldvegadvadidpnynqllhnnals 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNAVYGDQOGNNVATIIIVHMKNGSLV 360
DB 301 svlmgshngiepekvslllyggngpgarhdwnatvgykdqggnvatiiivhmknsgslv 360
QY 361 IAGGEGINNPFSYLYKEDQLTGSORALSQEIQNKIDFMFLAONNAKLNDLSKEKEK 420
DB 361 iaggeginnpfsylykedqltgsoralseqeiqnkidfmeflaonnakldnlsekekek 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAUSKDKTHSALITEFGNGDLSYTLKDYGGKADKA 480
DB 421 frteikdfqkdsakayldalgnndriaufavskdkthsalitefgngdlsytlkdygkka 480
QY 481 LDREKNVTLOGLSKHDGVNMFVDYSNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAIFNL 540
DB 481 ldreknvtlqgslkhdgvnmfvdyasnfytnasknppkvgvgtngvshlevgnkvaifnl 540
QY 541 PDLNNLAITSFVRRLNEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTTLNFKAVADAKN 600
DB 541 pdlnnlaitsfvrnlndeklttkglspqeaanklikdfissnkelygttlfnfkavadakn 600
QY 601 TGNDEVKKAQKLEKSLRKHLEKEVEKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
DB 601 tgndevkkaqkleslsrkhelekevekklesksgnknkmeakagansqkdeifalin 660
QY 661 KEANDARAIAYAOQLKIGIKRELSDKLENVKNKLDFOKSPDFPKNGKNDKSKAEETLK 720
DB 661 keandaraiaayaglnkigikrelsdklenvknkldfoksfdefkngkndkskaeetlk 720
QY 721 ALKGSVKDLGNPEWISKVENLNAALNEFKNGKNDKSKVTQAKSDLENSVKDVIINOKV 780
DB 721 alkgsvkdlgnpewiskvenlnaalnefngkngkdfskvtqaksdlenstvdiingkv 780
QY 781 TDKVDNLNQAVSAKATGDFSRVEQALADLNQISGLVVGQAGPPLKRHDKVDLDSKVLGR 960
DB 781 tdkvdnlngavsvakatgdfsrveqaladlnqisglvvgqagpplkrhdkvdlksvglr 960
QY 841 KNGVNTLVNGLSQAEATTLSKNFSDIKKELNAKLNFNNNNNNNGLKNEPIYAKVKKK 900
DB 841 kngvntlvnglsqaeattlsknfsdikkelnaki nfnnnnnnglnknepiyavkvvkk 900
QY 901 AGQAAASLEPIYAQVAKKVNKIDRLNQTASGLVVGQAGPPLKRHDKVDLDSKVLGR 960
DB 901 agqaasleepiyaqvakkvnakidrlngiasglvvgqagpplkrhdkvdlksvglr 960
QY 961 NOELAQKIDNLNQAVSEAKAGFFGNLEQITDKLKDSTKHNPMLNWSAKKVPASLSAKL 1020
DB 961 nqelaqkidnlngavseakagffgnleqitdklkdstkhnpmnlwvesakkvpsalsakl 1020
QY 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLVDNKIVAHNVGSPVSEYDKIG 1080
DB 1021 dnyatnshirinsnikngaineekatgmilqknpewklvndkivahnvgsvpseydkig 1080
QY 1081 FNOKNMKDYSDFKFTKLNNAVKDTNSGFTQFLTNFSTASYCYCLARENASHGKKNVT 1140
DB 1081 fnoknmkdysdfkftklnnavkdtngftqfltnafstasycyclarenashgknkvt 1140
QY 1141 KGGFQKS 1147
DB 1141 kggfqs 1147
RESULT 3
AAR53269
ID AAR53269 standard; Protein; 1181 AA.
XX AAR53269;
XX AAR53269;

DT 13-DEC-1994 (first entry)
 XX Tag A antigen of Helicobacter pylori.
 DE Tag A; antigen; peptic ulceration; Helicobacter pylori; antibody;
 XX detection; diagnosis; therapy; treatment.
 KW Helicobacter pylori.
 XX OS
 XX WO9409023-A.
 XX PD 28-APR-1994.
 XX PF 13-OCT-1993; 93WO-US09782.
 XX PR 13-OCT-1992; 92US-0959940.
 XX (UYVA-) UNIV VANDERBILT.
 XX Blaser MJ, Cover TL, Tummuru MKR;
 XX WPI; 1994-151235/18.
 DR N-PSDB; AAQ64581.
 XX
 PT DNA coding tag A gene, from Helicobacter pylori - useful for
 PT detecting predisposition to peptic ulceration
 XX Claim 29; Page 68-72; 87pp; English.
 XX Monoclonal antibodies directed against the antigen may be used to
 CC detect tag A antigen presence which is indicative of a predisposition
 CC to peptic ulceration. A ligand e.g. Antibody, specifically reactive
 CC with the tag A antigen can be used to treat peptic ulcers.
 XX Sequence 1181 AA;
 SQ

Query Match 90.0%; Score 5281; DB 15; Length 1181;
 Best Local Similarity 88.4%; Pred. No. 1.8e-264;
 Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

QY 1 MTNETIDQOPQEAFAFPQOFINNLOVAFKVDNAVASVDPDKPIVDKNDNRNQAFEG 60
 DB 1 mnetidqppqceafnppqffinnlqvafkvdnavasvdpdkpivdkndnrnraqef 60
 QY 61 ISQLREYSNKAIAKPTKKNQYFSDFINKNDLINDVESSTKSFQFGDQRYRIF 120
 DB 61 isqlreysnkaiaikptkknqyfsdfinkndlindvesstksfkgdgyrif 120
 QY 121 TSWSVSHQNDPSKINTRSIRNFWENIQQPILDDKEAEFLKSAKQSPAGIIGNOIRTDQ 180
 DB 121 tswwshqndpskintsrirnfwenioppilddkeaeeflksakqspagiignoirtdq 180
 QY 181 KFMGVFDESILKEROBAEKNGEPTGDWLDIFLSFTFDKQSSDVKEAINQSPVHPVOPDI 240
 DB 181 kfmgvfdesilkerbaekngptgdwdiflsfthdkqssdvkeainqspvhpvpdi 240
 QY 241 ATTTTIDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
 DB 241 atstthigglpessrdlldergnfskftlgdmemldvegvadidpnykfnollhnnals 300
 QY 301 SVLMGSHNGIEPEKVSLLYGGNGPGARHDNATVGYKDOGNVNTIIVHMKNGSLV 360
 DB 301 svlmgshngiepekvslllyggngpgarhdnatvgykdognnvntiivhmkngslv 360
 QY 361 IAGGEGKINNPFSFVLYKEDQLTGSORALSQEEIQNKIDFMEFLAQNNAKLDNLSEKEKEK 420
 DB 361 iaggegkinnpfsfylvkyedqltgsoralsqeeiqnkidfmeflaqqnnakldnlsekeke 420
 QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAIVSKDKTKHSALITEFGNGDLSYTLKDYGKKADKA 480
 DB 421 frteikdfqkdsayldalgnndriaivskdktkhsalitefgngdlsytlkvmgkkaika 480

QY 481 LDREKNVTLOGSLKHGCVNMFVDYSNPKYTNASKNPNKGVGVTVNGVSHLEVGFNKAIFNL 540
 DB 481 ldreknvtlqgnlkhdgvmfvnysnfkytnaskspnkgvgtngvshleagfskvavfnl 540
 QY 541 PDLNNLAITSFVRNLEDKLTTKGLSPQEAANKLRDQFLSSNKLKCTLNFNKAVADAKN 600
 DB 541 pldnnlaitsfvrrnledklttkglspqeaanklrvdflssnkelvgkalnfkavaean 600
 QY 601 TGNVDEVKKAQDKLEKSLKREHLEKEVEKLESNGKNKMEAKAQAQNSQKDEIFALIN 660
 DB 601 tgnvdevkraqdleklkrehlekdvaknlesngknkmeakaqaqnsqkdeifallin 660
 QY 661 KEANDARAIAYAQNLKLGIRKRELSKLENVNKNLKDQKDFDFKNGKNDKFSKABETLK 720
 DB 661 keanrdaraiayaqnlkgikrelsdkleninkldkfskdfgkngknkdfksaeetlk 720
 QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVDVVIINQKV 780
 DB 721 alkgsvkdilginpewiskvenlnaalnefkngknkdfskvtqaksdqlensikdviiinqkl 780
 QY 781 TDKVDNLNQAQSVAKATGDFSRVQALADLKNFSKEOLAQAQAKNESLNARKKSEIYQSV 840
 DB 781 tdkvdelnqavsvakiadfsveqaladlknfskeqlaqgqaknesfnv-gkselygsv 839
 QY 841 KNGVNGTLVGNLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNGLKN--EPIYAKVNMK 898
 DB 840 kngvngtlvnglsgieatalaknfidkkelneknfnfnnnnnnglknnggepiyavnmk 899
 QY 899 KKAGAAASLEPIYAQVAKKVNNAKIDRLNQIA-SGLGVVVGQA----- 939
 DB 900 kktgvaspeepiyavakvktkidqlnqaatsfgvggagfplkrhdkvedlskvr 959
 QY 940 -----AGPPLKRHDKVDLIDLVKGLSRNQLAOKIDNLNQAQSEAKAGFEG 984
 DB 940 svspepiyatiddlggsfplkrhdkvdldskvlsrneltdqldnlsqavseakagffg 1019
 QY 985 NLEQITDKLDKSTKHNPMLNWSAKVPASLSAKLDNATNNSHIRINSNIKNGAINKA 1044
 DB 1020 nleqitdkldkftknpvnlwaesakvpsasakldnatsnshirinsnlnqngaineka 1079
 QY 1045 TGMLTOKNPEWLKLVNDKIVAHNVGSLSEYDKTGFNOKNMKYSDSKFSTKLNNNAVK 1104
 DB 1080 tgmtoqnpewlkivndkivahnvgsplseydnigfsqknmkdysdskfstklnnnavk 1139
 QY 1105 DTNSGFTQFLNAPFASVYVCLARENAEHGKKNVNTKGFQKS 1147
 DB 1140 diksgftqlanafast-gyysmarenaehgkknantkgygfdks 1181

RESULT 4
 AAR72593
 ID AAR72593 standard; Protein; 1181 AA.
 XX AC AAR72593;
 XX DT 29-SEP-1995 (first entry)
 XX H. pylori tagA antigen.
 XX TagA; antigen; ulcer; diagnosis; vaccine.
 XX Helicobacter pylori.
 XX US5403924-A.
 XX PD 04-APR-1995.
 XX PF 13-OCT-1992; 92US-0959940.
 XX PR 13-OCT-1992; 92US-0959940.
 XX PR 26-APR-1993; 93US-0053614.
 XX (UYVA-) UNIV VANDERBILT.
 PA

XX PI Blaser MJ, Cover TL, Tummuru MKR;
 XX WPI: 1995-146855/19.
 DR N-PSDB; AAQ86728.
 XX
 PT New nucleic acid encoding tag A antigen of Helicobacter pylori -
 PT used to detect predisposition to peptic ulceration and to produce
 PT protein for use in vaccines, diagnosis etc.
 XX
 PS Disclosure; Column 37-46; 30pp; English.
 XX
 CC The full-length sequence of the tagA gene of H. pylori 84-183 (ATCC
 53726) was obtained from overlapping clones isolated from genomic
 CC libraries. The gene encoded a 1181-amino acid TagA antigen
 CC protein.
 XX
 SQ Sequence 1181 AA;

Query Match 90.0%; Score 5281; DB 16; Length 1181;
 Best Local Similarity 88.4%; Pred. No. 1.8e-264;
 Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

QY 1 MTNETIDQQPQTEAFAFPQPFINNLQVAFILKVDNAVASYDDPQKPIVDKNDNRQAFEG 60
 DB 1 mtnetidqqpqtteafnpgqfnnlqvafilkvdnavasyddpdkpivdkndnrqafeg 60
 QY 61 ISQLEREYSNKAIKNPYKNOYFDFNKNNDLNLKNDLIDVESSTKSFQKFGQRYRIF 120
 DB 61 isqleereysnkaiKNPYKNOYFDFNKNNDLNLKNDLIDVESSTKSFQKFGQRYRIF 120
 QY 121 TSWYSHQNDPSKINTRSIRNFEMENIIQPPILDDKEKAEFLKSAKQSFAGIIIGNQIRTDQ 180
 DB 121 tswyshqndpskintrsirnfmennioppilddkkekafelksakqsfagiignqirtdq 180
 QY 181 KFMGVFDESILKEROEAKNGEPTGGDWLIDFLSIFDKKQSSDVKEAINGEPVHVQPD 240
 DB 181 kfmgvfdesilkerqeaekngptggdwldflsifdkkqssdvkeaingeplhvqpd 240
 QY 241 ATTTTDTQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLJHNNALS 300
 DB 241 atstthtqglppesrdlldergnfskftlgdmemldveg vadidpnykfnnlljhnntls 300
 QY 301 SVLMGSHNGIEPERKVSLLYGGNGPGARHDMNATVGYKDQGNVATIIIVHMKNGSLV 360
 DB 301 svlmgshngieperkvsllayagngfgakhwnatvgykdqgnvatiivhmkngsglv 360
 QY 361 IAGGEGINNPFSYLYREDQDTGSORALSQEEIQNKIDFMFEFLAONNAKLNLSEKEKEK 420
 DB 361 iaggeginnpsfilykedqdtgsqralsqeeiqnkidfmeflaqnnaakldslsekekek 420
 QY 421 FRTEKDFQDKSKAYLDALGNDRTAFVSKKDTKHSALITFEGNGDLSYTLKDYCKKADKA 480
 DB 421 frnelkdfqdkskayldalgndrtafvskkdpkhsalitefnkdlsytlkvmmkqkika 480
 QY 481 LDREKNVTLOGSLKHGDMVFDYSNFKYTNAKKNPNKPGVGTNGVSHLEVGFNKVAIFNL 540
 DB 481 ldreknvtlogslkhgdmvfnysnfkynaskspnkgvgtngvshleagfsvavfnl 540
 QY 541 PDLNLAITSFVRNLEDKLTGKLSPOEANKLTDFLSSNNKELVGKTLNENKAVADAKN 600
 DB 541 pldlnlaitsfvrrnledkltakglspqeaanklvkdfllssnkelvgkalnfnkavaeakn 600
 QY 601 TGNVDEYKKAQDKLESLKRREHLEKEVEKKLESKSGNKNKMEAKAANSOKDEIFALIN 660
 DB 601 tgnvdevkradqdkleslkrrehlekdvaknlesksgnknkmeakagansqkdelifalin 660
 QY 661 KEANRDAIRAIAQNLKIGIKRELSDKLENVKNLKDQKSFDEFKNGKNKDFSKAEETLK 720
 DB 661 keanrdairaiaqnlkigirelsdkleninkldkfskfdgkngknkdfskaeetlk 720
 QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780

DB 721 alkgsvkdilginpewiskvenlnaalnfnknknkdfskvtqaksdgnskidvinqki 780
 QY 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKEQALQAQAKNESLNARKKSEIYOSV 840
 DB 781 tdkvdelinqavsvakiacdsgveqaladlknfskeqlaqaqaknesfnv-gkseiyqsv 839
 QY 841 KNGVNGTLVNGLSQAEATTLTKNFSDIKKELNAKLGNFNNNNNGLKN--EPIYAKVKNK 898
 DB 840 kngvngtlvnglsgieatalaknfsdikkelnkfnfnnnnnnglknnggepiyaqvnk 899
 QY 899 KKAGQASLEPEPIYAQAVAKKVNNAKIDRLNOIA-SGLGVVGOA----- 939
 DB 900 kktggvaspeepiyaqavakvktkiddinqaatsgfgvgqagfplkrhdkvedlskvgr 959
 QY 940 -----AGFPLKRHRDKVDLskvplskrdkvdlskvrlnrqelqkldnlsqavseakgffg 1019
 DB 960 svspepiyatidldlgsgfplkrhdkvdlskvrlnrqelqkldnlsqavseakgffg 1019
 QY 985 NLEQFIDKLDKSTKHPNMLWVESAKKVPASLSAKLDNYATNHSIRINSNIKNGAINEKA 1044
 DB 1020 nleqtidklkdfktnpvnlnwaesakkvpaslsakldnyatnshtrinsninqaineke 1079
 QY 1045 TGMLTQKNPEWLKLVNDKIVAHNVGSPVLSYDKIGFNQKNMKDYSDSFKEFTKLNNNAVK 1104
 DB 1080 tgrtqknpewiklvndkivahnvgsplseydnigfsqkmmkdydsdfkfstklennavk 1139
 QY 1105 DTNSGTFQFTNFASTASYCLARENAEPHGKIKNTYTKGGFOKS 1147
 DB 1140 diksgftqlanaft-gyyismarenaehgiknantkggfoks 1181

RESULT 5
 AAR91307
 ID AAR91307 standard; Protein; 1181 AA.
 XX AAR91307;
 XX 09-SEP-1996 (first entry)
 XX Helicobacter pylori TagA 120-128 kD antigen.
 XX Antigen; peptic ulcer; chronic gastritis; gastric adenocarcinoma;
 KW diagnosis; predisposition; antibody; vaccine; infection.
 XX Helicobacter pylori.
 XX WO9610639-A2.
 XX 11-APR-1996.
 XX 29-SEP-1995; 95WO-US12669.
 XX 30-SEP-1994; 94US-0316397.
 XX (ORAV-) ORAVAX INC.
 XX (UYVA-) UNIV VANDERBILT.
 XX Blaser MJ, Cover TL, Kleanthous H, Tummuru MKR;
 XX WPI: 1996-209361/21.
 XX N-PSDB; AAT14051.
 XX Helicobacter pylori Tag A gene - used to develop prods. for the
 PT diagnosis, treatment and prevention of peptic ulceration and gastric
 PT carcinoma
 XX Claim 4; Page 103-107; 118pp; English.
 XX AAR91307 is a 120-128 kD antigen of Helicobacter pylori (HP). The
 CC antigen is designated TagA and it, or its fragments, can be used to
 CC determine a predisposition to peptic ulceration or gastric carcinoma,
 CC conditions caused by or linked to HP infection. TagA and its

CC fragments may also be used for antibody prodn. for use in detection
 CC of tagA in patients suspected of HP infection. TagA antibodies or
 CC other ligands may also be used to treat peptic ulceration or gastric
 CC carcinoma caused by HP infection. TagA or a non-functional TagA
 CC mutant may be used in vaccines for preventing and treating HP
 CC infection.
 XX
 SQ Sequence 1181 AA;

Query Match 90.0%; Score 5281; DB 17; Length 1181;
 Best Local Similarity 88.4%; Pred. No. 1.8e-264;
 Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

QY 1 MTNETIDQPOTEAFNPOQFINNQLQVAFKVDNNAVASYDDQKPIVDKNDNRNQAPEG 60
 DB 1 mtnetidqppqteafnpoqfinnqlqvfikvdnavasyddqkplvdndrnqrqafeg 60

QY 61 ISQLREYSNKAIKNPTTKNQVFSDFINKSNDLKNLIDVESSTKSQKFGDQRYTF 120
 DB 61 isqlreysnkaiknpttknqvfsdfinksndlnklndlivesstksqkfgdgryrif 120

QY 121 TSWSHQNDPSKINTRSIRNFENIIOPILODDKEAEFLKSAKOSFAGIITGNQIRTDQ 180
 DB 121 tswsHQNDPSKINTRCINFMHTCIQPPIDDKAEFLKSAKOSFAGIITGNQIRTDQ 180

QY 181 KFMGVDESLEKQEAENKGEPTGGDWLIDFLSFIDFKKQSSDVKEAINQEPVPHVQPD 240
 DB 181 kfmgvdeslekergeaenkgeptggdwidflsfidfkqssdvkeainqepvphvqpd 240

QY 241 ATTWTDIOGLPEARDLDERGNFSGFTLGDMEMLDVEGVADIDPNYFNQLLIHNNALS 300
 DB 241 atstthidgllpesrdldergnfsgftlgdmemldvegvadidpnynfqnllihnnals 300

QY 301 SVLMGSHNGIEPEKYSLYGGNGPGARDWNATVGYKDDQGNVATINVHMKNKSGSLV 360
 DB 301 svlmgshngiepekysllyggngpgardwnatvgykddqgnvatinvhmknksgslv 360

QY 361 IAGGEGKINNPFFYLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLNDLSEKEK 420
 DB 361 iaggeginnpfflykyedqltgsoralseqeieqknkidfmeflagnnakiidsekeke 420

QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAIVSKKTKHSALTERTGNGDLSTYTLKDYGRKAKA 480
 DB 421 frneikdfqkdkayldalgnndriaivsktkhsaltertgngdlsytlkdygrkaka 480

QY 481 LDREKNVTLOGLSKHGDGMVDFVSNFKVTNASKNKGVTNGVSHLEVGFKVAIFNL 540
 DB 481 ldreknvtlqgnlkhdgvmvdfvsnfkvtnasknkgvtnvgvshleagfskvavfnl 540

QY 541 PDLNLAITSFVRRLDKLTGKLSPOEANKLIKDFLSSNKELVGTINFNKAVADAKN 600
 DB 541 pldlnlaitsvvrldedkltgklspeanklkvdfllssnkelvgkalnfknavaeakn 600

QY 601 TCNYDEVKKAQKDLKSKLRKREHLEKEVEKKLESKSGNKNMEAKAQAANSOKDEIFALIN 660
 DB 601 tcnvdevkkaqkdlkskkrrehlekevekklesksgnknmeakagaansokdeifalin 660

QY 661 KEANDARAIAVAONLKGKIRELSDKLENVKNLKDFFKDFEFGKNGKDFSKAEETLK 720
 DB 661 keandaraiavaonlkgkirelsdklenvknlkdfdkdfefgkngknkdfskaeetlk 720

QY 721 ALKGSVKDLGINPEWISKVENLNAALNFKNGKNDKFSKVTQAQSDLENSVKDVIINGKV 780
 DB 721 alkgsvkdlginpewiskvenlnaalnfknkndkfskvtqaqsdleNSVKDVIINGKV 780

QY 781 TDKVDNLQAVSVAKATGDFSRVEQALADLNFSEKQLAQQAQKNESUNARKKSEIYQSV 840
 DB 781 tdkvdnlqavsvakatkdfsrveqaladlnfsekeqlaqqaqknesunv-gkseIYQSV 840

QY 841 KNGVNGTLVGNGLSQAEATTLKSNFSDIKKELNAKLGNFNNNNNNNGLKN--EPIYAKVKN 898
 DB 840 kngvngtlvgnglsqaeatTLKSNFSDIKKELNAKLGNFNNNNNNNGLKN--EPIYAKVKN 898

QY 899 KKAGQAASLEPIYAQVAKKVNAKIDRLNQIA-SGLGVVVGQA----- 939
 DB 900 kktgvaspeepiyaqvakvkckkldqlloqaatsfgvggagfplkrhdkvedlskvgr 959

QY 940 -----AGFPLKRHDKVDDLSKVGLSRNOELAQKIDNLNOAVSEAKAGFTG 984
 DB 960 svspepiyatidldlgsgfplkrhdkvddlskvglsrnqeltqkldhlsqavseakagf 1019

QY 985 NLEQITDKLKSTKHNPMNLVWESAKKVPASLSAKLDNATNSHINRINSKNGAINAKA 1044
 DB 1020 nleqitdklkdftknpmnlvawesakkvpsakldnatnshtrinsnigaina 1079

QY 1045 TQMLTQKNPEWMLKLVNDKIVAHNVGSPVSEYDKTGFNCKNMKDYSDSFKSTKLNNAYK 1104
 DB 1080 tqterqknpewlklvndkivahnvsgvplseydnigfsqkmmkdydsfksfcklennavk 1139

QY 1105 DFNSTGFTFLTNAFTASYCIARENAEHGKINVTAKGFGFKS 1147
 DB 1140 diksgftqlanafst-gyysmarenaehgknantkggfgks 1181

RESULT 6
 AAW20174
 ID AAW20174 standard; protein; 1178 AA.
 XX AC AAW20174;
 XX DT 08-JUL-1997 (first entry)
 XX DE H. pylori cytoplasmic protein, 16459375.aa.
 XX KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 XX OS Helicobacter pylori.
 XX PN WO9640893-A1.
 XX PD 19-DEC-1996.
 XX PF 06-JUN-1996; 96WO-US09122.
 XX PR 01-APR-1996; 96US-0630405.
 XX PR 07-JUN-1995; 95US-0487032.
 XX PA (ASTR) ASTRA AB.
 XX PI Berglindh OT, Smith D, Mellgaerd BL;
 XX DR WPI: 1997-052306/05.
 XX DR N-PSDB; AAT67410.
 XX PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX Claim 61; Pages 390-392; 1481pp; English.
 PS The present sequence is a Helicobacter pylori cytoplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori.
 CC infection or to identify H. pylori polypeptide binding compounds, useful
 CC as potential H. pylori life cycle activators or inhibitors. The genomic
 CC sequence of H. pylori (ATCC 35679) was determined from overlapping
 CC contigs generated by mechanically shearing the bacterial DNA. The
 CC sequences were analysed for ORF of at least 180 nucleotides, and the
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide

506 -----tepiyctqavkxvkakiorloqiasgigavgqaa511k1rndkvaadlskv 954
 507 -----

Query Match 86.7%; Score 5086; DB 18; Length 1183;
Best Local Similarity 84.1%; Pred. No. 2.1e-254;
Matches 996; Conservative 60; Mismatches 73; Indels 56; Gaps 4;

QY 1 MTNETIDQOPTEAAFPQOFINNQLQVAFKLVYDNAVASVDPQOKPIVDKNDNRDNRAQAFEG 60
DB 17 mnealngqptqteafnpgqfnnlqvafikvdyvavfapnqkplvdkndrdrnqafek 76
QY 61 ISQLEESYNAKAIKPTTKNQYFSDFINKSNDLINKNLIDVESYTKSFQFGDQRYRIF 120
DB 77 isqlreefankaikptkknqyfsfisksndldkdnldigtssiksfqkfgtqyqif 136
QY 121 TSWVSHQNDPSKINTSRINFMENIIQPPILDDKEAEFLKSAKOSFAGIIIGNOIRTDQ 180
DB 137 mhwshqndpskintqkirkfmenliqppisddkeaeftlsakqafagiignqirsdq 196
QY 181 KFMGVFDESLEKROEAKNGE----PTGGDWLIDFLSFDFKQSSDVKEAINQEPVPHV 236
DB 197 kfmgvfdeslkergeaekngpndgtgtdwldflsfdfkqssdlketlngpvpvhv 256
QY 237 QPDIATTTDIOGLPEARDLDDGNFSEKFTLGDWEMLDVBGVADIDPNYKFNOLLJHN 296
DB 257 qpdvatttdiqslppearldldernfskftlgdmmlldvegvdidpnykfnqllhm 316
QY 297 NALSSVLMGSHNGIEPEKVSLLYGGGPGARHDNATVGYKDOQGNVATIINVMKNG 356
DB 317 nalssvlmgshngiepekvslllygngpgpearhwnatvgykngqgdnvatlinvmkng 376
QY 357 SGLVITAGEKGINNPSFYLRKEDQLTGSORALSQSEIQNKIDFMEFLAQNNAKLDNSEK 416
DB 377 sgvlvtaggekgnpsfylyrkedqltgsqrlsqeiegnkvdfmeflaqnnaakldnlsk 436
QY 417 EKEKPERTEIKDPQDSKAYLDALGNDRIAFYSKDKTISALITERGNDLSVTLKDYCKK 476
DB 437 ekekpeiedfqskayldalgnndhrafvskdkkhlalvaeftngelsytlkdyckk 496
QY 477 ADKALDREKNTVLOGSLKHGDVYDFYNFYTNASKNPNGVGTNGVSHLEVGFNKVA 536
DB 497 adkaldreaktllgslkhgdvymfydfnfytnaskspdkvgvatngvshleagfskva 556
QY 537 IFNLPDLNLAITSFVRNLEDKLTTKGLSPOEANKLKDFLSSNKELVGTGLNPNKAVA 596
DB 557 vfnlplnlnlatsvvrqdeokliakglspqeanlkvdfllssnkkelvgkalnfkava 616
QY 597 DAKNTGNTDEVKKAQDKLESLRKREHLEKEVEKKLESKGNKNKMEAKAQAANSOKDEIF 656
DB 617 eakntgnydevkkaqdkleslkrerlekdvaknlesksgnknkmeaksgansqkdeif 676
QY 657 ALINKEANDRATAYANLKGIRKRELSKLENVNKNLDFDKSPDEFKNGKNKDFSKAE 716
DB 677 alinkeandratayaqnlkgirkrelsdkleninkdikfsksdefkngknkdfskae 736
QY 717 ETLKALGKSVKDLGINPEWISKVENLNAALNEFFNKGKKNKDFSKVTOAKSDLENSKVDVII 776
DB 737 etlkalgksvkdlginpewiskvenlnaalnefnkgknkdfskvtqaksdlsensikdvii 796
QY 777 NQKVTDKVDNLNOAVSAKATGDFSRVBOALADLKNFSKEOLAOQAKNESLNARKKSEI 836
DB 797 nqkicdkvdnlnoavsvakatydgfsgvveqaladlknfskeqlaqagkqneofntgknsai 856
QY 837 YQSVKNGVNGTLVGNGLSQAEATTLTSKNFSDIKKELNAKLGNNNNNNGLKNPEIYAKV 896
DB 857 yqsvkngvngtlvngnlskaeatltsknfsdikkelnaaklgnnnnnnnglens ----- 910
QY 897 NKKAGQAASLEPIYAOVAKVNAKIDRLNQIASGLGVGQOAG----- 941
DB 911 -----tepiytqavkvvakidrlqiasglgdvggaasfllkrhdkvdldiskv 959
QY 942 -----FPLKRHDKVDDLskvslsrnqelaokIDNLNOAVSEAKAGF 982
DB 960 glsanhepiyatidldlgpflkrhdkvdldiskvslsrnqelqtqkldnlnqavseakash 1019

QY 983 FGNLEQOTIDKLDSTKHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSNIKNGAINE 1042
DB 1020 fgnlqgidkldstckknvnlyvesakkvptlsakldnyatnshtrlnsnvknqgtine 1079
QY 1043 KATGMLTKQNPBWLKLVNDKIVAHNVGVSPVLPSEYDKIGFNQKNMKDYSDSFESTKLNA 1102
DB 1080 katgmiltgknsewlkvlndkivahnvgvaplsaydkigfnqknmkdysdfkstrlsna 1139
QY 1103 VKDTSNGFTQTLTNAFSTASYCLARENAEHGKNNVNTKGGFOKS 1147
DB 1140 vkdksgfvqfntlnfsmgs-yslmkasvehgvkntntkkgfqs 1183

RESULT 8

AAR53268
ID AAR53268 standard; Protein; 859 AA.

AC AAR53268;

DT 13-DEC-1994 (first entry)

XX 120-128 kilodalton antigen of Helicobacter pylori.

XX Tag A; antigen; peptic ulceration; Helicobacter pylori; antibody;
KW detection; diagnosis.

OS Helicobacter pylori.

PN W09409023-A.

XX 28-APR-1994.

XX 13-OCT-1993; 93WO-US09782.

XX 13-OCT-1992; 92US-0959940.

PA (UYVA-) UNIV VANDERBILT.

PI Blaser MJ, Cover TL, Tummuru MKR;

DR WPI; 1994-151235/18.

DR N-PSDB; AAQ64580.

PT DNA coding tag A gene, from Helicobacter pylori - useful for
detecting predisposition to peptic ulceration

XX Claim 3; Page 58-61; 87pp; English.

CC The antigenic fragment is a fragment of the tag A antigen of
Helicobacter pylori. Monoclonal antibodies directed against the
antigen may be used to detect tag A antigen presence which is
indicative of a predisposition to peptic ulceration.

XX Sequence 859 AA;

Query Match 69.1%; Score 4056.5; DB 15; Length 859;
Best Local Similarity 92.7%; Pred. No. 1.7e-201;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;

QY 1 MTNETIDQOPTEAAFPQOFINNQLQVAFKLVYDNAVASVDPQOKPIVDKNDNRDNRAQAFEG 60

DB 1 mtnetidqopqteafnpgqfnnlqvafikvdyvavfapnqkplvdkndrdrnqafeg 60

QY 61 ISQLEESYNAKAIKPTTKNQYFSDFINKSNDLINKNLIDVESYTKSFQFGDQRYRIF 120

DB 61 isqlreeysnkaiknptkknqyfsfisksndldkdnldigtssiksfqkfgtqyqif 120

QY 121 TSWVSHQNDPSKINTSRINFMENIIQPPILDDKEAEFLKSAKOSFAGIIIGNOIRTDQ 180

DB 121 tswvshqndpskintcrlrnfmehtiqpplddkkaeflksakosfagiignqirtdq 180

QY 181 KFMGVFDESLEKROEAKNGEPTGGDWLIDFLSFDFKQSSDVKEAINQEPVPHVQPD 240

```
Db 181 kmvgfdeslkerqaekngptggdwldflsfldkqssdvkeainqepiphvqpd 240
Qy 241 ATTTTIDIOGLPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 atstchigglpdesrdlldergnfskftlgdmemldvegvdmdpnkfnqllihntls 300
Qy 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNAATVGYKDDQGNVATIIIVHMKNGSLV 360
Db 301 svlmshgdgiepeksvlliyagngfgakhdwatvgvkdqggnvatiiinvhmknsglv 360
Qy 361 TAGGEGKINNPSFYLYKEDLTGSGORALSOEIQNKIDPFMEFLAONNAKLNLSEKEK 420
Db 361 taggegkinpsfylykedqltgsgralsqeeiqnkidpfmeqlagmnaklslsekeke 420
Qy 421 FRTEIKDFQKDSKAYLDALGNDRIAFVSKDKTPKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 frneikdfqkdkpyldalgnndriaafvskdktpkhsallitefngdlsytlkvmgkkika 480
Qy 481 LDREKNVTLOGSLKHGDMFVDFYSNFKYTNASKNPNKGVGTNGVSHLEVGNKVAIFNL 540
Db 481 ldreknvtlqgnlkhdgvmfvnysnfkytnaskspnkgvgtngvshleagfskvavfnl 540
Qy 541 PDLNNLATTSFVRRNLEDKLTTKGLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 pdlnnlatstvrrnledklttkglspoeanklikdflssnkelygktnlfnkavadeakn 600
Qy 601 TGNVDEVKKAQKDLKSLKRHEHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Db 601 tgnvdevkraqdlekslkrtehlekdvaknlesksnknkmeakaqansqkdeifalin 660
Qy 661 KEANDARAIAYAQNKLGIKRELSKLENVNKNLDFKSFDEPFNGKNKDFSKAEETLK 720
Db 661 keanararaiayaqnlkgikrelskleninkldkfskfdgfkngknkdfskaeetlk 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDLINQKV 780
Db 721 alkgsvkdlginpewiskenlnaalnefkngknkdfskvtqaksdqensikdvlingki 780
Qy 781 TDKVDNLNQAQVSAKATGDFSRVEQALADLNKFSKEQLAQAQKNESLNARKKSEIYQSV 840
Db 781 tdkvdelnqaqvsaiaedfsgveqaladlnkfskeqlaqaqknesfnv-gkseiyqsv 839
Qy 841 KNGVNGTLVGNGLSQAEAT 859
Db 840 kngvngtlvnglsgieat 858

RESULT 9
ID AAR72594 standard; Protein; 859 AA.
XX
AC AAR72594;
XX
DT 29-SEP-1995 (first entry)
XX
DE H. pylori tagA antigen (truncated).
XX
KW TagA; antigen; ulcer; diagnosis; vaccine.
XX
OS Helicobacter pylori.
XX
PN US5403924-A.
XX
PD 04-APR-1995.
XX
PF 13-OCT-1992; 92US-0959940.
XX
PR 13-OCT-1992; 92US-0959940.
XX
PR 26-APR-1993; 93US-0053614.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
```

```
PI Blaser MJ, Cover TL, Tummuru MKR;
DR WPI; 1995-146855/19.
DR N-PSDB; AAQ86728.
XX
PT New nucleic acid encoding tag A antigen of Helicobacter pylori -
PT protein for use in vaccines, diagnosis etc.
XX
PS Disclosure: Column 37-46; 30pp: English.
XX
CC The full-length sequence of the tagA gene of H. pylori 84-183 (ATCC
CC 53726) was obtained from overlapping clones isolated from genomic
CC libraries. The gene encoded a 1181-amino acid TagA antigen
CC protein (AAR72593) and a truncated antigen (AAR72594).
XX
SQ Sequence 859 AA;

Query Match 69.1%; Score 4056.5; DB 16; Length 859;
Best Local Similarity 92.7%; Pred. No. 1.7e-201;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;

Qy 1 MTNETIDQOPTEAAFPNQFINNLQVAFLLKYDNAVASVDPDQKPLVDKNDNRDNRQAPEG 60
Db 1 mtneidqgqpqteaafnpqfinnlqvafllkydnavaasydpdqkplvdkndnrdrnqafeg 60
Qy 61 ISQLEEYSNKAIKNPTKKNQYFSDFINKSNDLINDKNDLIDVESSTKSFQKFGDQYRIF 120
Db 61 isqleeysnkaiknptkknqyfsdfinksndlindkndlivvesstksfqkfgdqyrif 120
Qy 121 TSWVSHQNDPSKINPSTRINFMENIIQPIILDKKAEPFLSKSAKSFAGIITGNIQRTDQ 180
Db 121 tswvshqndpskintrcirnmehitqppildkkaeflksaksfagiitgningirtq 180
Qy 181 KFMGVFDESLEKEROEAKNGEPTGGDWLDIFLSFTFDKQSSDVKEATNOEVPVHVOPDI 240
Db 181 kfmgvfdeslkerqaekngptggdwldiflsftfdkqssdvkeainoevpvhvopdi 240
Qy 241 ATTTTIDIOGLPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 atstchigglpdesrdlldergnfskftlgdmemldvegvdmdpnkfnqllihntls 300
Qy 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNAATVGYKDDQGNVATIIIVHMKNGSLV 360
Db 301 svlmshgdgiepeksvlliyagngfgakhdwatvgvkdqggnvatiiinvhmknsglv 360
Qy 361 TAGGEGKINNPSFYLYKEDLTGSGORALSOEIQNKIDPFMEFLAONNAKLNLSEKEK 420
Db 361 taggegkinpsfylykedqltgsgralsqeeiqnkidpfmeqlagmnaklslsekeke 420
Qy 421 FRTEIKDFQKDSKAYLDALGNDRIAFVSKDKTPKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 frneikdfqkdkpyldalgnndriaafvskdktpkhsallitefngdlsytlkvmgkkika 480
Qy 481 LDREKNVTLOGSLKHGDMFVDFYSNFKYTNASKNPNKGVGTNGVSHLEVGNKVAIFNL 540
Db 481 ldreknvtlqgnlkhdgvmfvnysnfkytnaskspnkgvgtngvshleagfskvavfnl 540
Qy 541 PDLNNLATTSFVRRNLEDKLTTKGLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 pdlnnlatstvrrnledklttkglspoeanklikdflssnkelygktnlfnkavadeakn 600
Qy 601 TGNVDEVKKAQKDLKSLKRHEHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Db 601 tgnvdevkraqdlekslkrtehlekdvaknlesksnknkmeakaqansqkdeifalin 660
Qy 661 KEANDARAIAYAQNKLGIKRELSKLENVNKNLDFKSFDEPFNGKNKDFSKAEETLK 720
Db 661 keanararaiayaqnlkgikrelskleninkldkfskfdgfkngknkdfskaeetlk 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDLINQKV 780
Db 721 alkgsvkdlginpewiskenlnaalnefkngknkdfskvtqaksdqensikdvlingki 780
```

Db 721 alkgsavldglnpewiskvenlnaalnefngknkdfskvtqaksdgsikdvlingki 780

Qy 781 TDVNDLNQAVSVAKATGDFSRVEQALADLNKFSKEQLAQQAQAKNESLNARKKSEIYQSV 840

Db 781 tdkvdelnqavsvakiacdsgveqaladlnkfskeqlaqaqaknesfnv-gkseiyqsv 839

Qy 841 KNGVNGTLVGNCLSOAEAT 859

Db 840 kngvngtlvnglsgleat 858

RESULT 10

AAR91308

ID AAR91308 standard; Protein: 850 AA.

XX

AC AAR91308;

XX

DT 09-SEP-1996 (first entry)

DE Helicobacter pylori TagA antigen 96 kD antigenic fragment.

XX

KW Antigen: peptic ulcer; chronic gastritis; gastric adenocarcinoma;

KW diagnosis; predisposition; antibody; vaccine; infection.

XX

OS Helicobacter pylori.

XX

PN WO9610639-A2.

XX

PD 11-APR-1996.

XX

PF 29-SEP-1995; 95WO-US12669.

XX

PR 30-SEP-1994; 94US-0316397.

XX

PA (ORAV-) ORAVAX INC.

PA (UYVA-) UNIV VANDERBILT.

XX

PI Blaser MJ, Cover TL, Kleanthous H, Tummuru MKR;

XX

DR WPI; 1996-209361/21.

DR N-PSDB; AAT14052.

XX

PT Helicobacter pylori Tag A gene - used to develop prods. for the

PT diagnosis, treatment and prevention of peptic ulceration and gastric

PT carcinoma

XX

PS Claim 5; Page 93-96; 118pp; English.

XX

CC AAR91308 is a 96 kD antigenic fragment of the Helicobacter pylori (HP)

CC TagA antigen. TagA, or its fragments, can be used to determine a

CC predisposition to peptic ulceration or gastric carcinoma, both

CC conditions caused by or linked to HP infection. TagA and its

CC fragments may also be used for antibody prodn. for use in detection

CC of TagA in patients suspected of HP infection. TagA antibodies or

CC other ligands may also be used to treat peptic ulceration or gastric

CC carcinoma caused by HP infection. TagA or a non-functional TagA

CC mutant may be used in vaccines for preventing and treating HP

CC infection.

XX

SQ Sequence 850 AA;

Query Match 68.7%; Score 4031.5; DB 17; Length 850;

Best Local Similarity 92.8%; Pred. No. 3.3e-200;

Matches 790; Conservative 24; Mismatches 36; Indels 1; Gaps 1;

Qy 1 MTNETIDQPOPEAAFPQQTINNLOVAFKVDNAVASYDDPKPIVDKNDNRNQAPEG 60

Db 1 mtnetidqppqteaaafnpqqfnnlqvafkvdnavasyddpkpivdkndrdnraqefg 60

Qy 61 ISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDLNLDVESSTKSFQKFGDQRYRIF 120

Db 61 isqlreeysnkaiknptkknqyfsdfinksnldlinkdnlivvssstksfkgfgdgyrif 120

Qy 121 TSWSHQNDPSKINTRSIRNFEMENIQQPIILDREKAEFLSAKQSFAGIIGNQIRTDQ 180

Db 121 tswshqndpskintrcirnfemehitqppipddkekaeflksaksfagilignqirtdq 180

Qy 181 KFMGVFDESLSKERQPAEKNGEPTGGDWLDIFLSFTFKQSSDYKAEALNQEPVPHVQPD 240

Db 181 kfmgvfdeslkergeaknggptggdwldiflsftfkqssdvkealinqepvphvqpd 240

Qy 241 ATTTTDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYFNOLLINNNALS 300

Db 241 atstthdgioppesrdllidergnfskftlgdmemldvegvadmdpnynkfngllhnnltls 300

Qy 301 SVLMGSHNGIEPEKVSLLYGGNGPGFARGHMDNATVGYKDDQGNVATIIINHMKNGSLV 360

Db 301 svlmgshdgiepekvsllyagngvgfakhdwnatvgykddqgnnvatiinhvmkngsglv 360

Qy 361 TAGGKGINNPSFYLYKEDQLTGSGRALSQEIQNKIDFMEFLAQNNAKLNLSEKEK 420

Db 361 taggkginpnsfylykedqitgsqralsqeigknkldfmeflaqqnakldslsekeke 420

Qy 421 FRTEIKDFQKDSKAYLDALGNDRIAFVSKKDTKHSALITEFGNGDLSYTLKDYGGKAKDA 480

Db 421 frnelkdfqkdsipyldalgnrdriafvskkdpkhsallitefnkgdlsylclvmgkqika 480

Qy 481 LDREKNVTLOGSLKHGDMFVDYSNFKYTNASKNPKNGVGTNGVSHLEVGFKNVAIFNL 540

Db 481 ldreknvtlqgnlkhdgvmfvnysnfkytnaskspkngvgtngvshleagfsvavfnl 540

Qy 541 PDLNNLAITSFVRRLNEDKLTGKLSPOEANKLIKDFLSSNNKELVGKTLNFMKAVADAKN 600

Db 541 pldnnlaitsvvrldedkllakglspqeanklvkdfissnkelygkalnfnkavaeakn 600

Qy 601 TGNVDEVKAKODLEKSLRKRHLEKEVEKKLESKSGKNKMEAKAQANSOKDETFALIN 660

Db 601 tgnvdevkradqkdekslkrhlekdvaknlesksgknkmeakaqansqkdeifalin 660

Qy 661 KEANDARAIAYAQNKGKIKRELSKLENVANKLNKDFDKSFDEPKNGKNKDFSKAEETIK 720

Db 661 keandaraiaayaqnkgkikrelsdkleninkdkfaksfdgfkngknkdfskaeetlk 720

Qy 721 ALKGSVKDLGINPEWISKVENLNALNEFKNGKNKDFSKVTQAKSDLENVSKDVIIINOKV 780

Db 721 alkgsvkdlginpewiskvenlnaalnefngknkdfskvtqaksdgsikdvlingki 780

Qy 781 TDVNDLNQAVSVAKATGDFSRVEQALADLNKFSKEQLAQQAQAKNESLNARKKSEIYQSV 840

Db 781 tdkvdelnqavsvakiacdsgveqaladlnkfskeqlaqaqaknesfnv-gkseiyqsv 839

Qy 841 KNGVNGTLVGN 851

Db 840 kngvngtlvgn 850

RESULT 11

AAW89997

ID AAW89997 standard; Protein: 318 AA.

XX

AC AAW89997;

XX

DT 18-FEB-1999 (first entry)

DE Expressed antigen for cluster 20.

XX

KW Antigen: immunogenic cluster family; vaccine; gastritis; diagnosis;

KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.

XX

OS Helicobacter pylori.

XX

PN WO9849314-A2.

XX

PD 05-NOV-1998.

XX


```

PF 27-APR-1998; 98WO-US08487.
XX
XX 14-OCT-1997; 97US-0061958.
XX 25-APR-1997; 97US-0045107.
XX
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX
XX Chow TP, Fry KE, Lim MY, McAtee CP;
XX
XX WPI; 1999-009433/01.
XX
XX New Helicobacter pylori antigens and related nucleic acid sequences
XX - useful in serological diagnosis and protective vaccines, providing
XX long-lasting immune response
XX
XX Claim 16; Page 329; 402pp; English.
XX
XX The present sequence represents a Helicobacter pylori antigenic protein
XX that is characterised by immunoreactivity with H. pylori-positive
XX antisera. The proteins are highly immunogenic and induce a long-lasting
XX immune response that persists even after antimicrobial treatment. In
XX antibody-detection assays, on sera, plasma, urine, saliva etc., they are
XX highly sensitive and specific. The specification also describes 69
XX previously unrecognised immunogenic cluster families. H. pylori antigens
XX are used to detect H. pylori-specific antibodies, for diagnosing
XX infection or to confirm eradication of infection, and in vaccines to
XX protect against H. pylori infection and related diseases (gastritis,
XX peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
XX Sequence 318 AA;

Query Match 27.1%; Score 1591; DB 20; Length 318;
Best Local Similarity 95.9%; Pred. No. 7.4e-75;
Matches 305; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 84 SDFINKNDLINKNDLIDVESSTKSFQFGDQRYDIFTWSVSHQNDPSKINTRSRNFM 143
Db 1 sdfidskndlinkndlidvestsksfqfgdrgyqifgtsvshqkdpksintrsrnfm 60

QY 144 NIIQPIIDDKKAEFLKSAQSAGIIGNOIRTDQFMGVFDESLEKROEAEKNGEPT 203
Db 61 niqppiddkkaeflksaqsagiiignoirtdqfmgvfdeslkergeaeknggpt 120

QY 204 GGDWLDIFLSFTFDKQSSDKVEATNQEPVHVQPIATTTTIDIGLPPEARDLLDERGN 263
Db 121 ggdwldiflsfinkkssdvkealngqepvhvqpdiattttidiglppeardlldern 180

QY 264 FSKFTLGDMEMLDVEGVADIDPNYKFNOLLTHNNALSSVLMGSHNGIEPEKVSLLYGNG 323
Db 181 fskftlgdmemldvegvadidpnykfnqllhnnalssvlgshngiepekvsllyagng 240

QY 324 GPGARHDNATVGYKDQGNVATTIIVHMKNGSLVITAGGEKGINNPFSFYLYKEDQLTG 383
Db 241 gfgdkhdnativgykdqgnvattinvhmkngslvliaggekginnpfsflykedqltg 300

QY 384 SQRALSQEEIQNKIDFME 401
Db 301 sqralsqeeirnkvdme 318

RESULT 12
AAW89878
ID AAW89878 standard; Protein: 187 AA.
XX
XX AAW89878;
XX
XX 18-FEB-1999 (first entry)
XX
XX Antigen 1 from cluster 18a.
XX
XX Antigen: immunogenic cluster family; vaccine; gastritis; diagnosis;
XX peptic ulcer; gastric adenocarcinoma; gastric lymphoma.

```

```

XX OS Helicobacter pylori.
XX PN WO9849314-A2.
XX PD 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08487.
XX
XX 14-OCT-1997; 97US-0061958.
XX 25-APR-1997; 97US-0045107.
XX
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX
XX Chow TP, Fry KE, Lim MY, McAtee CP;
XX
XX WPI; 1999-009433/01.
XX
XX New Helicobacter pylori antigens and related nucleic acid sequences
XX - useful in serological diagnosis and protective vaccines, providing
XX long-lasting immune response
XX
XX Claim 1; Page 225-226; 402pp; English.
XX
XX The present sequence represents a Helicobacter pylori antigenic protein
XX that is characterised by immunoreactivity with H. pylori-positive
XX antisera. The proteins are highly immunogenic and induce a long-lasting
XX immune response that persists even after antimicrobial treatment. In
XX antibody-detection assays, on sera, plasma, urine, saliva etc., they are
XX highly sensitive and specific. The specification also describes 69
XX previously unrecognised immunogenic cluster families. H. pylori antigens
XX are used to detect H. pylori-specific antibodies, for diagnosing
XX infection or to confirm eradication of infection, and in vaccines to
XX protect against H. pylori infection and related diseases (gastritis,
XX peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
XX Sequence 187 AA;

Query Match 15.6%; Score 917; DB 20; Length 187;
Best Local Similarity 94.7%; Pred. No. 2.3e-40;
Matches 177; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTNETIDQOPQTEAARNFQFQFINNLOVAFKVDNAVSYDPPQKPIVDKNDNRDNRQAFEG 60
Db 1 mtnetidqpgqteaaafnpqfnnlqvafkvdnavasydpqkplvdnkndrnrqafdg 60

QY 61 ISOLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKNDLIDVESSTKSFQFGDQRYTF 120
Db 61 ssqireeysnkaiknptkknqyfsdfinesndlinkndlidgssiksfgkfgtqyrif 120

QY 121 TSWVSHQNDPSKINTRSRNFMENIQQPILDDKKAFLKSAQSAGIITGNQIRTDQ 180
Db 121 tswwshqndpskintrsrnfmenniqqpildkkaeflksaqsagiiitgnqirtdq 180

QY 181 KFMGVFD 187
Db 181 kfmgvfd 187

RESULT 13
AAG82935
ID AAG82935 standard; Protein: 5024 AA.
XX
XX AAG82935;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis open reading frame protein sequence SEQ ID NO:2964.
XX
XX Staphylococcus epidermidis SR1 strain; infection; diagnosis;
XX vaccination; endocarditis.

```


RESULT 15
AAW24575
ID AAW24575
XX


```

Qy 1017 SAKLDNYATNSH-----IRINSIKNGAINEKATGMLTQKNPEWKLIVNDKIVAH 1066
Db 1045 seeltkavtdsktiisklgvlievnentemntiessakeiealyn-----elknkktsln 1100
Qy 1067 NV---GSVPLSEYDKIGFNQKNMKDYSDSKFSTKLNNAVKDYNSTGFTQPLTNAFSTAS 1122
Db 1101 eiqtstnevkige-----mksnadkyldvskifntvldtqk--snivtn----- 1142
Qy 1123 YYCLARENAEHGKNVNTK 1141
Db 1143 -----ghsinnvkdk 1152

```

Search completed: February 13, 2002, 14:00:58
Job time: 3507 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2002, 13:05:31 ; Search time 62.29 Seconds
(without alignments)
414.373 Million cell updates/sec

Title: US-09-360-685A-5
Perfect score: 5867
Sequence: 1 WYNEIDQOPQTEAAFNPOQ.....RENAEHGKNVTKGFGQKS 1147

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5867	100.0	1147	3	US-08-470-260-5
2	5867	100.0	1147	3	US-08-471-491-5
3	5867	100.0	1147	4	US-08-466-662-5
4	5867	100.0	3289	2	US-08-477-451-2
5	5281	90.0	1181	1	US-08-053-614-4
6	5281	90.0	1181	1	US-08-316-397B-4
7	5281	90.0	1181	2	US-09-034-306-4
8	5281	90.0	1181	4	US-09-259-437-4
9	5281	90.0	1181	5	PCT-US93-09782-4
10	4056.5	69.1	859	1	US-08-033-614-2
11	4056.5	69.1	859	1	US-08-316-397B-2
12	4056.5	69.1	859	2	US-09-034-306-2
13	4056.5	69.1	859	4	US-09-259-437-2
14	4056.5	69.1	859	5	PCT-US93-09782-2
15	284	4.8	2285	4	US-09-308-375-2
16	248.5	4.2	1312	2	US-08-592-126-148
17	248.5	4.2	1312	2	US-08-687-080-51
18	246	4.2	1354	3	US-08-685-871-2
19	238.5	4.1	2710	1	US-08-480-604A-6
20	238.5	4.1	2710	2	US-08-405-496A-6
21	238.5	4.1	2710	4	US-08-915-136-6
22	232	4.0	1085	1	US-08-431-080-28
23	232	4.0	1085	2	US-08-938-534-28
24	231	3.9	1535	3	US-08-755-587-185
25	231	3.9	2391	2	US-08-446-855A-2
26	231	3.9	2391	4	US-09-150-741-2
27	230	3.9	1588	5	PCT-US93-07261-11

28	230	3.9	1663	5	PCT-US93-07261-16
29	228	3.9	1600	2	US-08-617-697-10
30	227.5	3.9	2482	1	US-08-328-254-6
31	226.5	3.9	1786	4	US-08-973-462-8
32	226.5	3.9	3248	1	US-08-353-700-1
33	226.5	3.9	3248	5	PCT-US95-16216-1
34	225.5	3.8	885	2	US-08-533-306A-4
35	225.5	3.8	885	2	US-08-742-923A-4
36	221	3.8	1529	2	US-08-728-470-10
37	221	3.8	1529	4	US-08-719-641-10
38	220	3.7	1664	2	US-08-642-846-2
39	216	3.7	816	2	US-08-533-306A-6
40	216	3.7	816	2	US-08-742-923A-6
41	215	3.7	1494	3	US-08-755-587-186
42	213	3.6	1579	3	US-08-755-587-184
43	212	3.6	1346	1	US-08-471-033-23
44	212	3.6	1346	2	US-08-471-044-23
45	212	3.6	1346	2	US-08-463-483A-23

ALIGNMENTS

RESULT 1
US-08-470-260-5
; Sequence 5, Application US/08470260
; Patent No. 6077706
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
; TITLE OF INVENTION: for Vaccines and Diagnostics
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,848
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1147 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-470-260-5

Query Match 100.0% Score 5867; DB 3; Length 1147;

```
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTNETIDQOPQTEAAFNPOQFINNLOVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOPQTEAAFNPOQFINNLOVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
Qy 61 ISQLEEYSNKAINKPTTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFQFGDQRYRIF 120
Db 61 ISQLEEYSNKAINKPTTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFQFGDQRYRIF 120
Qy 121 TSWVSHQNDPSKINTRSIRNFEMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHQNDPSKINTRSIRNFEMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Qy 181 KFMGVFDESLEKROEAEKNGEPTGGDWLDIFLSFTFDKQSSDVKEAINQEPVHVQDPI 240
Db 181 KFMGVFDESLEKROEAEKNGEPTGGDWLDIFLSFTFDKQSSDVKEAINQEPVHVQDPI 240
Qy 241 ATTTTIDQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYFNQOLLIHNNALS 300
Db 241 ATTTTIDQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYFNQOLLIHNNALS 300
Qy 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKDOQGNVATTINVHMKNKSGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKDOQGNVATTINVHMKNKSGSLV 360
Qy 361 TAGGEGKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLONLSEKEK 420
Db 361 TAGGEGKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLONLSEKEK 420
Qy 421 FRTEIKDFQKDSKAYLDALGNDRIFAFAVSKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 FRTEIKDFQKDSKAYLDALGNDRIFAFAVSKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480
Qy 481 LDREKNVTLQGSKLHGDGMVFDVSNFKYTNASKPNKGVGTNGVSHLEVGFKNKVAIFNL 540
Db 481 LDREKNVTLQGSKLHGDGMVFDVSNFKYTNASKPNKGVGTNGVSHLEVGFKNKVAIFNL 540
Qy 541 PDLNNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKAVADAKN 600
Db 541 PDLNNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKAVADAKN 600
Qy 601 TCNYDEVKKAQDLKSLRKRHLKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
Db 601 TCNYDEVKKAQDLKSLRKRHLKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
Qy 661 KEANRDARAIYAQNLKGIKRELSKLENVKNLKDPEKSFDEPKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIYAQNLKGIKRELSKLENVKNLKDPEKSFDEPKNGKNKDFSKAEETLK 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSYKDVITINOKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSYKDVITINOKV 780
Qy 781 TDKVDNLNAQVSAKATGDFSRVEQALADLNKFSKEQLAQQAQNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNAQVSAKATGDFSRVEQALADLNKFSKEQLAQQAQNESLNARKKSEIYQSV 840
Qy 841 KNGVNTLVGNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKK 900
Db 841 KNGVNTLVGNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKK 900
Qy 901 AGQAASLEPIYAQVAKVNAKIDRLNGIASGLGVGQAAGFPPLKRHRDKVDLKSGLSLR 960
Db 901 AGQAASLEPIYAQVAKVNAKIDRLNGIASGLGVGQAAGFPPLKRHRDKVDLKSGLSLR 960
Qy 961 NOELAQKIDNLNQAVSEAKAGFFGNLEQIDKLDKSTKHNPMLNWLWESAKVPASLSAKL 1020
Db 961 NOELAQKIDNLNQAVSEAKAGFFGNLEQIDKLDKSTKHNPMLNWLWESAKVPASLSAKL 1020
Qy 1021 DNYATNSHIRNSINKNAINEKATGMLTOKNPEWKLVDKIVAHNVGSPFLSEYDKIG 1080
Db 1021 DNYATNSHIRNSINKNAINEKATGMLTOKNPEWKLVDKIVAHNVGSPFLSEYDKIG 1080

RESULT 2
US-08-471-491-5
; Sequence 5, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0044
; CURRENT APPLICATION NUMBER: US/08/471.491B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-471-491-5

Query Match 100.0%; Score 5867; DB 3; Length 1147;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTNETIDQOPQTEAAFNPOQFINNLOVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOPQTEAAFNPOQFINNLOVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
Qy 61 ISQLEEYSNKAINKPTTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFQFGDQRYRIF 120
Db 61 ISQLEEYSNKAINKPTTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFQFGDQRYRIF 120
Qy 121 TSWVSHQNDPSKINTRSIRNFEMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHQNDPSKINTRSIRNFEMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Qy 181 KFMGVFDESLEKROEAEKNGEPTGGDWLDIFLSFTFDKQSSDVKEAINQEPVHVQDPI 240
Db 181 KFMGVFDESLEKROEAEKNGEPTGGDWLDIFLSFTFDKQSSDVKEAINQEPVHVQDPI 240
Qy 241 ATTTTIDQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYFNQOLLIHNNALS 300
Db 241 ATTTTIDQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYFNQOLLIHNNALS 300
Qy 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKDOQGNVATTINVHMKNKSGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKDOQGNVATTINVHMKNKSGSLV 360
Qy 361 TAGGEGKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLONLSEKEK 420
Db 361 TAGGEGKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLONLSEKEK 420
Qy 421 FRTEIKDFQKDSKAYLDALGNDRIFAFAVSKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 FRTEIKDFQKDSKAYLDALGNDRIFAFAVSKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480
Qy 481 LDREKNVTLQGSKLHGDGMVFDVSNFKYTNASKPNKGVGTNGVSHLEVGFKNKVAIFNL 540
Db 481 LDREKNVTLQGSKLHGDGMVFDVSNFKYTNASKPNKGVGTNGVSHLEVGFKNKVAIFNL 540
```



```
Db 481 LDREKNVTLOGSLKHGDMFVDFYDNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAIFNL 540
Qy 541 PDNLNLAITSEVRRLNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 PDNLNLAITSEVRRLNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Qy 601 TGNVDEVKKAQDKLESLRKREHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
Db 601 TGNVDEVKKAQDKLESLRKREHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
Qy 661 KEANRDARAIAAQNKGKIKRELSDKLENVKNLKDQDKSDFDEPKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIAAQNKGKIKRELSDKLENVKNLKDQDKSDFDEPKNGKNKDFSKAEETLK 720
Qy 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Qy 781 TDKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKEQLAQAQAKNESLNARKKSEIYOSV 840
Db 781 TDKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKEQLAQAQAKNESLNARKKSEIYOSV 840
Qy 841 KNGVNTLVGNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Db 841 KNGVNTLVGNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Qy 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVGQAAGFPPLKRHDVDDL SKVGLSR 960
Db 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVGQAAGFPPLKRHDVDDL SKVGLSR 960
Qy 961 NOELAQKIDNLNOAVSEAKAGFFGNLEQTDKLDKSTKHPMNLWVESAKKVPASLSAKL 1020
Db 961 NOELAQKIDNLNOAVSEAKAGFFGNLEQTDKLDKSTKHPMNLWVESAKKVPASLSAKL 1020
Qy 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLVDNKIVAHNVGVSPLSEYDKIG 1080
Db 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLVDNKIVAHNVGVSPLSEYDKIG 1080
```

RESULT 3

```
; Sequence 5, Application US/0846662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466,662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Helicobacter pylori
us-08-466-662-5
```

Query Match 100.0%; Score 5867; DB 4; Length 1147;
Best Local Similarity 100.0%; Pred. No. 0;

```
Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MTNETIDQOPQTEAAFNPOQFINNLQVAFKVDNAVAYDQDKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOPQTEAAFNPOQFINNLQVAFKVDNAVAYDQDKPIVDKNDNRDNRQAFEG 60
Qy 61 ISQREESYNAIKNPTKKNQYFSDFINKSNDLIDVESSTKSFQKFGDQRYRIF 120
Db 61 ISQREESYNAIKNPTKKNQYFSDFINKSNDLIDVESSTKSFQKFGDQRYRIF 120
Qy 121 TSWYSHQNDPSKINTSRIRNFEMENIQQPILDDKEKAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWYSHQNDPSKINTSRIRNFEMENIQQPILDDKEKAEFLKSAKOSFAGIIGNOIRTDQ 180
Qy 181 KFMGVDESLSKERQEAENKGEPTGGDWLIDFLSFIFDKKQSSDVKEAINOEPVPHVQDPI 240
Db 181 KFMGVDESLSKERQEAENKGEPTGGDWLIDFLSFIFDKKQSSDVKEAINOEPVPHVQDPI 240
Qy 241 ATTTTDDIOGLPPEARDLLDERGNFSKFTLGDMEMLDVGCVADIDPNYKFNQLLIHNNALS 300
Db 241 ATTTTDDIOGLPPEARDLLDERGNFSKFTLGDMEMLDVGCVADIDPNYKFNQLLIHNNALS 300
Qy 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKDOQGNVATTIINVHMKNGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKDOQGNVATTIINVHMKNGSLV 360
Qy 361 TAGKEGINNPSFYLYKEDQLTGSORALSQEEIQNKIDFMEFLAQNNAKLNDLSEKEKEK 420
Db 361 TAGKEGINNPSFYLYKEDQLTGSORALSQEEIQNKIDFMEFLAQNNAKLNDLSEKEKEK 420
Qy 421 FRTEIKDQKOSKAYLDALGNDRATFVSKKDTKHSALLITEFGNGDLSYTLKDYGGKADKA 480
Db 421 FRTEIKDQKOSKAYLDALGNDRATFVSKKDTKHSALLITEFGNGDLSYTLKDYGGKADKA 480
Qy 481 LDREKNVTLOGSLKHGDMFVDFYDNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMFVDFYDNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAIFNL 540
Qy 541 PDNLNLAITSEVRRLNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 PDNLNLAITSEVRRLNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Qy 601 TGNVDEVKKAQDKLESLRKREHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
Db 601 TGNVDEVKKAQDKLESLRKREHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
Qy 661 KEANRDARAIAAQNKGKIKRELSDKLENVKNLKDQDKSDFDEPKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIAAQNKGKIKRELSDKLENVKNLKDQDKSDFDEPKNGKNKDFSKAEETLK 720
Qy 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Qy 781 TDKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKEQLAQAQAKNESLNARKKSEIYOSV 840
Db 781 TDKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKEQLAQAQAKNESLNARKKSEIYOSV 840
Qy 841 KNGVNTLVGNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Db 841 KNGVNTLVGNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Qy 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVGQAAGFPPLKRHDVDDL SKVGLSR 960
Db 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVGQAAGFPPLKRHDVDDL SKVGLSR 960
Qy 961 NOELAQKIDNLNOAVSEAKAGFFGNLEQTDKLDKSTKHPMNLWVESAKKVPASLSAKL 1020
Db 961 NOELAQKIDNLNOAVSEAKAGFFGNLEQTDKLDKSTKHPMNLWVESAKKVPASLSAKL 1020
Qy 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLVDNKIVAHNVGVSPLSEYDKIG 1080
Db 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLVDNKIVAHNVGVSPLSEYDKIG 1080
```

QY 1081 FNQNMKDYSDSPKSTKLNNNAVDKNSGFTQFLTNAPSTASYCLARENAEHGKKNVT 1140
 Db 1081 FNQNMKDYSDSPKSTKLNNNAVDKNSGFTQFLTNAPSTASYCLARENAEHGKKNVT 1140
 QY 1141 KGGFQKS 1147
 Db 1141 KGGFQKS 1147
 RESULT 4
 US-08-477-451-2
 ; Sequence 2, Application US/08477451
 ; Patent No. 5928865
 ; GENERAL INFORMATION:
 ; APPLICANT: Covacci, Antonello
 ; TITLE OF INVENTION: Helicobacter Pylori CagI Region
 ; NUMBER OF SEQUENCES: 46
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,451
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McClung, Barbara G.
 ; REGISTRATION NUMBER: 33,113
 ; REFERENCE/DOCKET NUMBER: 0335.002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 510-601-2708
 ; TELEFAX: 510-655-3542
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3289 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-477-451-2

Query Match 100.0%; Score 5867; DB 2; Length 3289;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTNETIDQOQTEAFNPOQFINNQLQVAFKVDNNAVSYDDQKPIVDKNDNRNRQAFEG 60
 Db 1532 MTNETIDQOQTEAFNPOQFINNQLQVAFKVDNNAVSYDDQKPIVDKNDNRNRQAFEG 1591
 QY 61 ISQREEYSNKAIRNPTKKNQYFSDFINKSNLDLNKNDLIDVESSTKSFQKFGQDQRYRIF 120
 Db 1592 ISQREEYSNKAIRNPTKKNQYFSDFINKSNLDLNKNDLIDVESSTKSFQKFGQDQRYRIF 1651
 QY 121 TSWVSHQNDPSKINTRSRNFMENIQQPILDDKEAEFLKSQKSAFAGIIIGNQIRTDQ 180
 Db 1652 TSWVSHQNDPSKINTRSRNFMENIQQPILDDKEAEFLKSQKSAFAGIIIGNQIRTDQ 1711
 QY 181 KFMGVFDESLEKQEAENGFTGGDWDIDFLSFIDFKKQSSDYKEAINQEPVPHVQDPI 240
 Db 1712 KFMGVFDESLEKQEAENGFTGGDWDIDFLSFIDFKKQSSDYKEAINQEPVPHVQDPI 1771
 QY 241 ATTTTIDIGLPPARDLLDERGNFSKFTLGDMWMLDVGVDADIDPNYKFNOLLIHNNALS 300
 Db 241 ATTTTIDIGLPPARDLLDERGNFSKFTLGDMWMLDVGVDADIDPNYKFNOLLIHNNALS 300

Db 1772 ATTTTIDIGLPPARDLLDERGNFSKFTLGDMWMLDVGVDADIDPNYKFNOLLIHNNALS 1831
 QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWATVGYKDQGGNNVATIIINVHMKNGSGLV 360
 Db 1832 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWATVGYKDQGGNNVATIIINVHMKNGSGLV 1891
 QY 361 IAGGEGKINNPSFYLYKEDQLTGSORALSQEIEQNKIDPMEFLAQNNAKLDNLSEKEKEK 420
 Db 1892 IAGGEGKINNPSFYLYKEDQLTGSORALSQEIEQNKIDPMEFLAQNNAKLDNLSEKEKEK 1951
 QY 421 FRTEIKDFOKDSKAYLDALGNDRIAFVSKDTHKHSALITEFNGNDLSYTLKDYGGKADKA 480
 Db 1952 FRTEIKDFOKDSKAYLDALGNDRIAFVSKDTHKHSALITEFNGNDLSYTLKDYGGKADKA 2011
 QY 481 LDREKNVTLQGSLLKHGDMFVDYSNFKYTNASKNPKNGVGTNGVSHLEVGNKVAIFNL 540
 Db 2012 LDREKNVTLQGSLLKHGDMFVDYSNFKYTNASKNPKNGVGTNGVSHLEVGNKVAIFNL 2071
 QY 541 PDLNNLAITSFVRRNLEDKLTGKLSPOBANKLIKDFLSSNKELVGTGLNFNKAADAKN 600
 Db 2072 PDLNNLAITSFVRRNLEDKLTGKLSPOBANKLIKDFLSSNKELVGTGLNFNKAADAKN 2131
 QY 601 TGNVDEVKKAOKDLEKSLRKREHLEKEVEKKLESKSGNKNKMEAKAQAQNSQDEIFALIN 560
 Db 2132 TGNVDEVKKAOKDLEKSLRKREHLEKEVEKKLESKSGNKNKMEAKAQAQNSQDEIFALIN 2191
 QY 661 KEANDARAIAQAQNLKGIKRELSKLENNVKNLKDFOKSPDEFKNGKNKDFSKAEETUK 720
 Db 2192 KEANDARAIAQAQNLKGIKRELSKLENNVKNLKDFOKSPDEFKNGKNKDFSKAEETUK 2251
 QY 721 ALKGSVKDLGINPEWISKVENIQAALNEFFKNGKNKDFSKVTQAKSDLENSVKDVIINOKV 780
 Db 2252 ALKGSVKDLGINPEWISKVENIQAALNEFFKNGKNKDFSKVTQAKSDLENSVKDVIINOKV 2311
 QY 781 TDKVDNLNAQVSAKATGDFSRVEQALADLNKFSKEQLAQQAQAKNESLNARKKSEIYQSV 840
 Db 2312 TDKVDNLNAQVSAKATGDFSRVEQALADLNKFSKEQLAQQAQAKNESLNARKKSEIYQSV 2371
 QY 841 KNGVNGTLVGNLSQAEATTLKSNFSDIKKELNAKLGNFNNNNNNNNGLKNEPIYAKVNKK 900
 Db 2372 KNGVNGTLVGNLSQAEATTLKSNFSDIKKELNAKLGNFNNNNNNNNGLKNEPIYAKVNKK 2431
 QY 901 AGQAASLEPIYAQVAKKVNNAKIDRLNQTASGLGVGQAAGPPLKRHKVDVDDLKSVGLSR 960
 Db 2432 AGQAASLEPIYAQVAKKVNNAKIDRLNQTASGLGVGQAAGPPLKRHKVDVDDLKSVGLSR 2491
 QY 961 NOELAQKIDNLNOAVSEAKAGFFGNLEOTIDKLDKSTKHNPMLNVAESAKKVPASLSAKL 1020
 Db 2492 NOELAQKIDNLNOAVSEAKAGFFGNLEOTIDKLDKSTKHNPMLNVAESAKKVPASLSAKL 2551
 QY 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLKLVNDKIVAHNVGSPVPLSEYDKTG 1080
 Db 2552 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLKLVNDKIVAHNVGSPVPLSEYDKTG 2611
 QY 1081 FNQNMKDYSDSPKSTKLNNNAVDKNSGFTQFLTNAPSTASYCLARENAEHGKKNVT 1140
 Db 2612 FNQNMKDYSDSPKSTKLNNNAVDKNSGFTQFLTNAPSTASYCLARENAEHGKKNVT 2671
 QY 1141 KGGFQKS 1147
 Db 2672 KGGFQKS 2678

RESULT 5
 US-08-053-614-4
 ; Sequence 4, Application US/08053614
 ; Patent No. 5403924
 ; GENERAL INFORMATION:
 ; APPLICANT: COVER, TIMOTHY L.
 ; APPLICANT: BLASER, MARTIN J.
 ; APPLICANT: TUMMURU, MURALI K. R.
 ; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
 ; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/053,614
APPLICATION NUMBER: 19930426
FILING DATE: 19930426
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-053-614-4

Query Match 90.0%; Score 5281; DB 1; Length 1181;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

Qy 1 MTNETIDQPPQTEAFAFPQPPQINNQLQVAFKVDNAVASYDDPQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQPPQTEAFAFPQPPQINNQLQVAFKVDNAVASYDDPQKPIVDKNDNRDNRQAFEG 60

Qy 1 ISQUREEYSKAIKPNPKKNOYFSDFINKSNDLKNLNDLIVESTSKSFQKFGDQRYIF 120
Db 1 ISQUREEYSKAIKPNPKKNOYFSDFINKSNDLKNLNDLIVESTSKSFQKFGDQRYIF 120

Qy 121 TSWVSHQNDPSKINTSRINPMENIOPPIIDDKAEFLKSAKQSFAGIIGNOIRTDQ 180
Db 121 TSWVSHQNDPSKINTSRINPMENIOPPIIDDKAEFLKSAKQSFAGIIGNOIRTDQ 180

Qy 181 KFMGVFDESLEKQEAKEKNGEPTGGDWLIDFLSFIFDKKQSDVKEAINQEPVPHVQPD 240
Db 181 KFMGVFDESLEKQEAKEKNGEPTGGDWLIDFLSFIFDKKQSDVKEAINQEPVPHVQPD 240

Qy 241 ATTTTIDIGLPPEARDLLDERGNSKFTFLGDMELMDVEGVADIDPNYKFNOLLHNALS 300
Db 241 ATSTTHIQGLPESRDLLDERGNSKFTFLGDMELMDVEGVADIDPNYKFNOLLHNALS 300

Qy 301 SVLMSHGNGIEPEKVSLLYGCNGPGARHDWNAVYGVYKQDQGNVATINVMKNGSLV 360
Db 301 SVLMSHGNGIEPEKVSLLYACNGGEGAKHDWNAVYGVYKQDQGNVATINVMKNGSLV 360

Qy 361 IAGKEGINNPSFYLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLNLSEKEK 420
Db 361 IAGKEGINNPSFYLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLNLSEKEK 420

Qy 421 FRTEIKFQKSKAYLDALGNDRIFAYSKDKTSHALITEFGNGDLSYTLKDYKKA 480
Db 421 FRNEIKFQKSKAYLDALGNDRIFAYSKDKTSHALITEFGNGDLSYTLKVMGKKQ 480

Qy 481 LDREKNTYLOGSLKHGDMFYVDYFNFKYTNASKPNKNGVGTNGVSHLEVGFKNVAFNL 540
Db 481 LDREKNTYLOGSLKHGDMFYVDYFNFKYTNASKPNKNGVGTNGVSHLEVGFKNVAFNL 540

Qy 541 PDANNLAITSFVRNLEDKLTTKGLSPOEANKLKDFLSSNKELVYKTLNFKAVADAKN 600
Db 541 PDANNLAITSFVRNLEDKLTTKGLSPOEANKLKDFLSSNKELVYKTLNFKAVAEAKN 600

Qy 601 TGNVDEVKKAQDLEKSLRKRHELEKEVEKLEKSGNKNKMEAKAQAANSOKDEIFALIN 660
Db 601 TGNVDEVKKAQDLEKSLRKRHELEKEVEKLEKSGNKNKMEAKAQAANSOKDEIFALIN 660

Qy 661 KEANRDARAIAYAOQLKGIKRELSDKLENVKNLKDFOKSEDEFKNGKNKFOFSKAEETLK 720
Db 661 KEANRDARAIAYAOQLKGIKRELSDKLENVKNLKDFOKSEDEFKNGKNKFOFSKAEETLK 720

Qy 721 ALKGSVDKLGINPEWISKVENLNAALNEFKNGKNKFOFSKVTOAKSDLENSKVDLIINOKV 780
Db 721 ALKGSVDKLGINPEWISKVENLNAALNEFKNGKNKFOFSKVTOAKSDLENSKVDLIINOKV 780

Qy 781 TDKVDNLNOAVSAKATGDFSRVQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNOAVSAKATGDFSRVQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840

Qy 841 KNGVNGTLVGNGLSQAEATTLTKNFSDFDIKELNAKLGNFNNNNNGLKN--EPIYAKVYK 898
Db 841 KNGVNGTLVGNGLSGIEATATLAKNFSDIKELNEKFKFNFNNNNNGLKNGGEPYIAQVYK 899

Qy 899 KKAGQASLEPIYIAQVAKVNAKIDRLNOIA--SGLGVYVQA----- 939
Db 899 KKAGQASLEPIYIAQVAKVNAKIDRLNOIA--SGLGVYVQA----- 939

Qy 939 KKTGOVASPEEPIYIAQVAKVTKIDQLNQAAATSGFVGQAGFPLKRRHKVDLSKVG 959
Db 939 KKTGOVASPEEPIYIAQVAKVTKIDQLNQAAATSGFVGQAGFPLKRRHKVDLSKVG 959

Qy 940 -----AGFPLKRRHKVDLSKVGSLSRNQLAQKIDNLNOAVSEAKAGFFG 984
Db 940 -----AGFPLKRRHKVDLSKVGSLSRNQLAQKIDNLNOAVSEAKAGFFG 984

Qy 984 SVSPEPIYATIDDLGGSPFLARHKVDLSKVGSLSRNQLAQKIDNLNOAVSEAKAGFFG 1019
Db 984 SVSPEPIYATIDDLGGSPFLARHKVDLSKVGSLSRNQLAQKIDNLNOAVSEAKAGFFG 1019

Qy 1044 NLEQTIIDKLDSTKHNPMLNVAESAKVPASLSAKLDYATNHSIRINSNTKNGAINSKA 1044
Db 1044 NLEQTIIDKLDSTKHNPMLNVAESAKVPASLSAKLDYATNHSIRINSNTKNGAINSKA 1044

Qy 1080 TGTERTOKNPELKLVDKIVAHNVGSPVPLSEYDNIQKNNKDYSDSFKFTKLNNAYK 1139
Db 1080 TGTERTOKNPELKLVDKIVAHNVGSPVPLSEYDNIQKNNKDYSDSFKFTKLNNAYK 1139

Qy 1105 DTNSGTOFLNASTASYCLARENAEHGKNNVTGKGFQKS 1147
Db 1105 DTNSGTOFLNASTASYCLARENAEHGKNNVTGKGFQKS 1147

Qy 1140 DIKSGTOFLNAST--GYISWARENAEHGKNNVTGKGFQKS 1181
Db 1140 DIKSGTOFLNAST--GYISWARENAEHGKNNVTGKGFQKS 1181

RESULT 6
US-08-316-397B-4
; Sequence 4, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMMURU, MURALI K. R.
; TITLE OF INVENTION: THE Laga GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397B
; FILING DATE:
; CLASSIFICATION: 435

```

; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-316-397B-4

Query Match          90.0%; Score 5281; DB 1; Length 1181;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

QY 1 MTNETIDQOQTEAAFNQQFNNLQVAFKVDNVAASYDDPKPIVDKNDNRDNRQAFEG 60
DB 1 MTNETIDQOQTEAAFNQQFNNLQVAFKVDNVAASYDDPKPIVDKNDNRDNRQAFEG 60
QY 61 ISQLEEYSNKAIKNPTKKNQYFSDFINKSNDLINKNLDIVESSTKSFQKPGDQRYIF 120
DB 61 ISQLEEYSNKAIKNPTKKNQYFSDFINKSNDLINKNLDIVESSTKSFQKPGDQRYIF 120
QY 121 TSWVSHONDPSKINTSRNFMENITIOPIILDDEKAEFLKSAKOSFAGIIGNOIRTDQ 180
DB 121 TSWVSHONDPSKINTSRNFMENITIOPIILDDEKAEFLKSAKOSFAGIIGNOIRTDQ 180
QY 181 KPMGVFDESLEKQBAEKNGEPTGGDWLIDFLSFIFDKKQSDVKAEINQEPVHPQDPI 240
DB 181 KPMGVFDESLEKQBAEKNGEPTGGDWLIDFLSFIFDKKQSDVKAEINQEPVHPQDPI 240
QY 241 ATTTTIDIGLPEARDLLDERGNFSKFTLGDMEMLDVGVADIDPNYFNQLLIHNNALS 300
DB 241 ATTTTIDIGLPEARDLLDERGNFSKFTLGDMEMLDVGVADIDPNYFNQLLIHNNALS 300
QY 301 SVLMGSHNGIEPKVSLYLGNGGPGARHWNATVGYKDDQGNVATIIIVHMKNGSLV 360
DB 301 SVLMGSHNGIEPKVSLYLGNGGPGARHWNATVGYKDDQGNVATIIIVHMKNGSLV 360
QY 361 TAGGEGKINNPSFYLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLDNLSEKEK 420
DB 361 TAGGEGKINNPSFYLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLDNLSEKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIAFVSKDKTKHSALITERGNGDLSTYTKDYGKKADKA 480
DB 421 FRNEIKDFQKDSKAYLDALGNDRIAFVSKDKTKHSALITERGNGDLSTYTKDYGKKADKA 480
QY 481 LDREKNVTLOGSLKHGVMFYVSNFYKNTASNKNKGVGTNGVSHLEVGFKNKVAIFNL 540
DB 481 LDREKNVTLOGSLKHGVMFYVSNFYKNTASNKNKGVGTNGVSHLEVGFKNKVAIFNL 540
QY 541 PDLNNLAITSVYRRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVYKGTINFNKAVADAKN 600
DB 541 PNLNNLAITSVYRRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVYKGTINFNKAVADAKN 600
QY 601 TGNVDEVKKAQKDLKSLRKEHLEKEVEKKLESKGNKNKMEAKAQAANSOKDEIFALIN 660
DB 601 TGNVDEVKKAQKDLKSLRKEHLEKEVEKKLESKGNKNKMEAKAQAANSOKDEIFALIN 660
QY 661 KEANDARAIAAQNKGKIKRELSDKLENVANKLKDPEKSFDEPKNGKNKDFSKAEETLK 720
DB 661 KEANDARAIAAQNKGKIKRELSDKLENVANKLKDPEKSFDEPKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNALNEFKNGKNKDESKVTQAQSDLENSVKDVIINQKV 780
DB 721 ALKGSVKDLGINPEWISKVENLNALNEFKNGKNKDESKVTQAQSDLENSVKDVIINQKV 780
QY 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLNKFSKEQLAQQAQKNESLNARKKSEIYQSV 840

```

```

DB 781 TDKVDNLNQAVSVAKIACDFSGVEQALADLNKFSKEQLAQQAQKNESFNV-GKSEIYQSV 839
QY 841 KNGVNGTLVGNLSQAEATTLTSKNFSDIKKELNAKLGNFNNNNNCLKN--EPIYAKVVK 898
DB 840 KNGVNGTLVGNLSGTEATALAKNFSDIKKELNEFKNFNNNNNGLKNGGPEPIYAOVVK 899
QY 899 KKAGQAASLEEPIYAOVAKKVNAKIDRLNQIA-SGLGVVVGQA----- 939
DB 900 KKTGVASPEEPIYAOVAKKVTKKIDQLNQATSGFGVGQAGFPLKRHDKVEDLSKVCR 959
QY 940 -----AGFPLKRHDKVDLSDKVLGRNOELAQKIDNLNQAVSEAKAGFTG 984
DB 960 SVSPEPIYATIDDLGGSPFLKRHDKVDLSDKVLGRNOELTQKIDNLQAVSEAKAGFTG 1019
QY 985 NLEQITIDKLDKSTKHPNVLWVESAKKVPASLSAKLDNYATNSHIRINSINIKNGAINKA 1044
DB 1020 NLEQITIDKLDKSTKHPNVLWVESAKKVPASLSAKLDNYATNSHIRINSINIKNGAINKA 1079
QY 1045 TGMLTQKNPEWLKLVNDKIVAHNVGSPVLPSEYDKIGFNOKNMKYSDSFKFTKLNNAYK 1104
DB 1080 TGTGRQKNPEWLKLVNDKIVAHNVGSPVLPSEYDNIIGFSQKNMKDYSDSFKFTKLNNAYK 1139
QY 1105 DTNSGFTQPLTNFSTASYCLARENAEHGIKNVNTKGGFQKS 1147
DB 1140 DIKSGFTQFLANAFST-GYYSMAENAEHGIKNANTKGGFQKS 1181

```

```

RESULT 7
US-09-034-306-4
; Sequence 4, Application US/09034306
; Patent No. 5876943
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/034.306
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/316.397
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-034-306-4

```

Query Match 90.0%; Score 5281; DB 2; Length 1181;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

Qy 1 MTNETIDQDPQTEAAFPQOQFINNLQVAFKLVDAVSDPQKPIVDKNDNRQAFEG 60
Db 1 MTNETIDQDPQTEAAFPQOQFINNLQVAFKLVDAVSDPQKPIVDKNDNRQAFEG 60

Qy 61 ISQUREYSNKAINKNPTKKNOYFSDFTINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120
Db 61 ISQUREYSNKAINKNPTKKNOYFSDFTINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120

Qy 121 TSWVSHONDPSKINTRSIRNFMEINIOPTILDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHONDPSKINTRSIRNFMEIIOPTIPDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180

Qy 181 KFMGVFDESILKERQAEKNGEPTGGDWLDIFLSFIPDKKQSSDVKEAINQEPVPHVOPDI 240
Db 181 KFMGVFDESILKERQAEKNGEPTGGDWLDIFLSFIPDKKQSSDVKEAINQEPVPHVOPDI 240

Qy 241 ATTTTIDQGLPPARDLDRGNFSKFTLGMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
Db 241 ATSTTHIQGLPPESRDLDLDRGNFSKFTLGMEMLDVEGVADIDPNYKFNOLLIHNNALS 300

Qy 301 SVLMGSHNGIEPEKVSLLYGGNGPGARHDMNATVGYKQOQGNVATILNVHMKNGSLV 360
Db 301 SVLMGSHDGEPEKVSLLYAGNGGFGAKHDMNATVGYKQOQGNVATILNVHMKNGSLV 360

Qy 361 IAGEKGINNPSFYLYKEDQLTGSORALSQEBEIQNKIDFMFLAONNAKLDNLSEKKEK 420
Db 361 IAGEKGINNPSFYLYKEDQLTGSORALSQEBEIQNKIDFMFLAONNAKLDNLSEKKEK 420

Qy 421 FRTEIKDFQKDSKAYLDALGNDRIFAFVSKKDTKHSALITBFGNDLSYTLKDYKQKADKA 480
Db 421 FRNEIKDFQKDSKPYLDALGNDRIFAFVSKKDPKHSALITBFGNDLSYTLKVMGKKQIKA 480

Qy 481 LDREKNTVLOGSLKHDGVMEFVDSNFXYTNASKNPNGVGTNGVSHLEVGFNKVAFNL 540
Db 481 LDREKNTVLOGSLKHDGVMEFVDSNFXYTNASKNPNGVGTNGVSHLEVGFNKVAFNL 540

Qy 541 PDLNLAITSFVRNRNLDKLTITGSLPOEANKLIKDFLSSNKELVGKTLNFKAVADAKN 600
Db 541 PDLNLAITSVVRNLDKLTITGSLPOEANKLIKDFLSSNKELVGKTLNFKAVADAKN 600

Qy 601 TGNVDEYKKAQKDLKSLRKEHLEKEVEKKLESKGNKNKMEAKAQAQNSQKDEIFALIN 660
Db 601 TGNVDEYKKAQKDLKSLRKEHLEKEVEKKLESKGNKNKMEAKAQAQNSQKDEIFALIN 660

Qy 661 KEANRDARATAYAQNLKIGIKRELSDKLENVKNLKDFFKDFPKNGKNDKDFSKAETLK 720
Db 661 KEANRDARATAYAQNLKIGIKRELSDKLENVKNLKDFFKDFPKNGKNDKDFSKAETLK 720

Qy 721 ALKGSVKDLGINPEWISKVENLAALNEFKNGKNKDFSKVTOAKSDLENSKVDVINOKV 780
Db 721 ALKGSVKDLGINPEWISKVENLAALNEFKNGKNKDFSKVTOAKSDLENSKVDVINOKV 780

Qy 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNESLNAFKKSEIYQSV 840
Db 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNESLNAFKKSEIYQSV 840

Qy 841 KNGVNGTLVNGLSQAEATTLNRSNFSDIKKELNAKLNFNNNNNNGLN--EPIYAKVKN 898
Db 840 KNGVNGTLVNGLSGIEATLAKNFSDIKKELNAKLNFNNNNNNGLN--EPIYAKVKN 899

Qy 899 KKAQOASLEPIYAQVAKVNAKIDRLNOIA-SGLGVWGA----- 939
Db 900 KKTGVASPEPIYAQVAKVNAKIDRLNOIA-SGLGVWGA----- 939

Qy 940 -----AGFPLKRHDKVDLDSKVLGRNQELAQKIDNLNQAVSEAKAGFFG 984
Db 940 -----AGFPLKRHDKVDLDSKVLGRNQELAQKIDNLNQAVSEAKAGFFG 984

Qy 960 SVSPEPIYATIDDLGGSFPLKRHDKVDLDSKVLGRNQELAQKIDNLNQAVSEAKAGFFG 1019
Db 960 SVSPEPIYATIDDLGGSFPLKRHDKVDLDSKVLGRNQELAQKIDNLNQAVSEAKAGFFG 1019

Qy 985 NLEOTIDKLKDKSHNPMNLWVESAKKVPASLSAKLDNATNSHIRINSNIKNGAINEKA 1044
Db 1020 NLEOTIDKLKDKFTKNPNVNLWVESAKKVPASLSAKLDNATNSHIRINSNIKNGAINEKA 1079

Qy 1045 TGMULTQKNPEWLKLVNDKIVAHNVGVSPLSEYDKIGFNQKNKDYSDSFSTKLNNAVK 1104
Db 1080 TGTREQKNPEWLKLVNDKIVAHNVGVSPLSEYDNIGFSQKNKDYSDSFSTKLNNAVK 1139

Qy 1105 DTNSGPTQFLTNAESTASYCYCLARENAEHGKIKNTYTKGFFOKS 1147
Db 1140 DIKSGPTQFLTNAEST-GYYSMARENAEHGKIKNTYTKGFFOKS 1181

RESULT 8
US-09-259-437-4
: Sequence 4, Application US/09259437
: Patent No. 6153390
: GENERAL INFORMATION:
: APPLICANT: COVER, TIMOTHY L.
: APPLICANT: BLASER, MARTIN J.
: APPLICANT: HARRY KLEANTHOS
: APPLICANT: TUMURU, MURALI K.R.
: TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NEEDLE & ROSENBERG, P.C. 1200
: STREET: 127 Peachtree Street, Suite 1200
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: US/09/259,437
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA: US/08/316,397
: APPLICATION NUMBER: US/08/316,397
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: SPRATT, GWENDOLYN D.
: REGISTRATION NUMBER: 36,016
: REFERENCE/DOCKET NUMBER: 2200.030
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404/688-0770
: TELEFAX: 404/688-9880
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1181 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-259-437-4

Query Match 90.0%; Score 5281; DB 4; Length 1181;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

Qy 1 MTNETIDQDPQTEAAFPQOQFINNLQVAFKLVDAVSDPQKPIVDKNDNRQAFEG 60
Db 1 MTNETIDQDPQTEAAFPQOQFINNLQVAFKLVDAVSDPQKPIVDKNDNRQAFEG 60

Qy 61 ISQUREYSNKAINKNPTKKNOYFSDFTINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120
Db 61 ISQUREYSNKAINKNPTKKNOYFSDFTINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120

```

QY 121 TSWVSHQNDPSKINTSRIRNFEMENIIPILDDKEKAEFLKSAKQSFAGIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTSRIRNFEMENIIPILDDKEKAEFLKSAKQSFAGIIGNQIRTDQ 180
QY 181 KFMGVFDESLEKQERAEKNGEPTGGDWLDFLSFTFDKQSSDVKEAINOEPPHVPQDI 240
Db 181 KFMGVFDESLEKQERAEKNGEPTGGDWLDFLSFTFDKQSSDVKEAINOEPPHVPQDI 240
QY 241 ATTTDIOGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
Db 241 ATSTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDNATVGYKDOQGNVATTINVHMKNKSGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDNATVGYKDOQGNVATTINVHMKNKSGSLV 360
QY 361 TAGGEGKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLNDLSEKEK 420
Db 361 TAGGEGKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLNDLSEKEK 420
QY 421 FRTEIKDFOKSKAYLDALGNDRIFAIVSKDTKHSALITEFGNGDLSYTLKDYKKADKA 480
Db 421 FRNETKDFOKSKAYLDALGNDRIFAIVSKDTKHSALITEFGNGDLSYTLKDYKKADKA 480
QY 481 LDREKNVTLOGSLKHGDMVFNYSFKYTNASKPNKGVGTNGVSHLEVGNKVAIENL 540
Db 481 LDREKNVTLOGSLKHGDMVFNYSFKYTNASKPNKGVGTNGVSHLEVGNKVAIENL 540
QY 541 PDLNNLAITSFVRRLNLEDKLTGKLSPOEANKLIKDFLSSNKKELVGTLNFNKAVADAKN 600
Db 541 PNLNNLAITSFVRRLNLEDKLTGKLSPOEANKLIKDFLSSNKKELVGTLNFNKAVADAKN 600
QY 601 TGNVDEVKKAQDKLESLURKREHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Db 601 TGNVDEVKKAQDKLESLURKREHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
QY 661 KEANRDAIAIYAONLGIKIBELSDKLBNVKNLKDPEKSFDEPKNGKNKDFSKAEETLK 720
Db 661 KEANRDAIAIYAONLGIKIBELSDKLBNVKNLKDPEKSFDEPKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNALNEFKNGKNKDFSKVTOAKSDLENSKVDVIINOKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNALNEFKNGKNKDFSKVTOAKSDLENSKVDVIINOKV 780
QY 781 TDKVDNLNQAQVSAKATGDFSRVFOALADLKNFSKEQLAQQAQAKNESLARKKSEIYQSV 840
Db 781 TDKVDNLNQAQVSAKATGDFSRVFOALADLKNFSKEQLAQQAQAKNESLARKKSEIYQSV 840
QY 841 KNGVNGTLVGNLGSQAETTLKSNFSDIKKELNAKLGNFNNNNNGLKN--EPIYAKVKN 898
Db 841 KNGVNGTLVGNLGSQAETTLKSNFSDIKKELNAKLGNFNNNNNGLKN--EPIYAKVKN 898
QY 899 KKAGQAASLEPIYAQVAKKYNKIDRLNQIA-SGLGVVGOA----- 939
Db 899 KKAGQAASLEPIYAQVAKKYNKIDRLNQIA-SGLGVVGOA----- 939
QY 900 KKTQOVASPEPIYAQVAKKVTYKIDQLNQIATSGFGVGQAGPLKRDHDKVEDLSKVGR 959
Db 900 KKTQOVASPEPIYAQVAKKVTYKIDQLNQIATSGFGVGQAGPLKRDHDKVEDLSKVGR 959
QY 940 -----AGFPLKRHDKVDLDSKVLGSLRNQELAQKIDNLNQAQVSEAKAGFFG 984
Db 940 -----AGFPLKRHDKVDLDSKVLGSLRNQELAQKIDNLNQAQVSEAKAGFFG 984
QY 985 NLEQITDKLDKSTHNPWNLWVESAKVPASLAKLDNYATNSHIRNSINIKGAINIKA 1044
Db 985 NLEQITDKLDKSTHNPWNLWVESAKVPASLAKLDNYATNSHIRNSINIKGAINIKA 1044
QY 1045 TGMLTQKNPEWLKLVNDKIYAHNVGVSPLSEYDKIGFNOKNMKDYSDSEKFSFKLNNNAVK 1104
Db 1045 TGMLTQKNPEWLKLVNDKIYAHNVGVSPLSEYDKIGFNOKNMKDYSDSEKFSFKLNNNAVK 1104
QY 1080 TGTQKQKNPEWLKLVNDKIYAHNVGVSPLSEYDNIGFYSQKNMKDYSDSEKFSFKLNNNAVK 1139
Db 1080 TGTQKQKNPEWLKLVNDKIYAHNVGVSPLSEYDNIGFYSQKNMKDYSDSEKFSFKLNNNAVK 1139
QY 1105 DTNSGFTQFUTNAFSTASYICLARENAEHGKNVNTKGGFQKS 1147
Db 1105 DTNSGFTQFUTNAFST--GYYSMARENAEHGKNVNTKGGFQKS 1181

```

```

RESULT 9
PCT-US93-09782-4
; Sequence 4, Application PC/TUS9309782
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/09782
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-09782-4

```

```

Query Match 90.0%; Score 5281; DB 5; Length 1181;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

QY 1 MTNETIDQOQTEAFNPQQFINNLOVAFKVDNNAVSYDDQKPIVDKNDNRNQAFEG 60
Db 1 MTNETIDQOQTEAFNPQQFINNLOVAFKVDNNAVSYDDQKPIVDKNDNRNQAFEG 60
QY 61 ISQREEYSNAIKNPTKKNQYFSDFINKSNLDLNKNDLIDVESSTKSFQKFGDQYRIF 120
Db 61 ISQREEYSNAIKNPTKKNQYFSDFINKSNLDLNKNDLIDVESSTKSFQKFGDQYRIF 120
QY 121 TSWVSHQNDPSKINTSRIRNFEMENIIPILDDKEKAEFLKSAKQSFAGIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTSRIRNFEMENIIPILDDKEKAEFLKSAKQSFAGIIGNQIRTDQ 180
QY 181 KFMGVFDESLEKQERAEKNGEPTGGDWLDFLSFTFDKQSSDVKEAINOEPPHVPQDI 240
Db 181 KFMGVFDESLEKQERAEKNGEPTGGDWLDFLSFTFDKQSSDVKEAINOEPPHVPQDI 240
QY 241 ATTTDIOGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
Db 241 ATSTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDNATVGYKDOQGNVATTINVHMKNKSGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDNATVGYKDOQGNVATTINVHMKNKSGSLV 360
QY 361 IAGGEKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLNDLSEKEK 420
Db 361 IAGGEKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLNDLSEKEK 420

```

QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAQVSKDKTKHSAITFPGNGDLSYTLKDYGKKA 480
DB 421 FRNEIKDFQKDSKAYLDALGNDRIFAQVSKDKTKHSAITFPGNGDLSYTLKDYGKKA 480
QY 481 LDREKNVTLOGLSKHDGVNMFVDSYFNKYNASKNPNKGVGTNGVSHLGVFNKVAIFNL 540
DB 481 LDREKNVTLOGLSKHDGVNMFVDSYFNKYNASKNPNKGVGTNGVSHLGVFNKVAIFNL 540
QY 541 PDNLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
DB 541 PDNLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
QY 601 TGNDEVKKAQDKLEKSLKRRLHLEKEVEKLEKSGNKNKMEAKAQAANSQKDEIFALIN 660
DB 601 TGNDEVKKAQDKLEKSLKRRLHLEKEVEKLEKSGNKNKMEAKAQAANSQKDEIFALIN 660
QY 661 KEANDARAIAAQNKGKRELSDKLENVKNLDFDKSPDEFKNGKNKDFSKAEETLK 720
DB 661 KEANDARAIAAQNKGKRELSDKLENVKNLDFDKSPDEFKNGKNKDFSKAEETLK 720
QY 721 ALKGVSKDLGINPEWISKVENIINAALNFKNGKNKDFSKVTQAKSDLENSVKDVIINQKV 780
DB 721 ALKGVSKDLGINPEWISKVENIINAALNFKNGKNKDFSKVTQAKSDLENSVKDVIINQKV 780
QY 781 TDKNVNLNOAVSVAKATGDFSRVEQALADLKNPSKEQLAAQAKNESLNARKKSETIYQSV 840
DB 781 TDKNVNLNOAVSVAKATGDFSRVEQALADLKNPSKEQLAAQAKNESLNARKKSETIYQSV 840
QY 841 KNGVNGTLVGNLGSOEATLTSKNFSDTKKELNAKLNFKNNNNCLKN--EPIYAKVKN 898
DB 841 KNGVNGTLVGNLGSOEATLTSKNFSDTKKELNAKLNFKNNNNCLKN--EPIYAKVKN 898
QY 899 KRAGQAASLEPIYAAQVAKVNAKIDRLNQIA--SGLGVVYQGA-----939
DB 899 KRAGQAASLEPIYAAQVAKVNAKIDRLNQIA--SGLGVVYQGA-----939
QY 900 KKTGOVASPEEPIYAAQVAKVTKIDQLNQAATSGFGVGQAGFPLKRHDKVEDLSKVGR 959
DB 900 KKTGOVASPEEPIYAAQVAKVTKIDQLNQAATSGFGVGQAGFPLKRHDKVEDLSKVGR 959
QY 940 -----AGPPLKRHDKVEDLSKVGLSRNQLAQKIDNLSQAVSEAKAGFFG 1019
DB 940 -----AGPPLKRHDKVEDLSKVGLSRNQLAQKIDNLSQAVSEAKAGFFG 1019
QY 985 NLEQTDIKLQSTKTPNMLWESAKKVPASLSAKLDNATVATNSHTRINSNIKNGAINNEKA 1044
DB 985 NLEQTDIKLQSTKTPNMLWESAKKVPASLSAKLDNATVATNSHTRINSNIKNGAINNEKA 1044
QY 1045 TGMLAQKNPEWELKLVNDKIVAHNVGSVPLSEYDKIGFNOKNMKDYSDSPKFTKLNNAVYK 1104
DB 1045 TGMLAQKNPEWELKLVNDKIVAHNVGSVPLSEYDKIGFNOKNMKDYSDSPKFTKLNNAVYK 1104
QY 1080 TGTQRQKNPEWELKLVNDKIVAHNVGSVPLSEYDNIQFSGKNMKDYSDSPKFTKLNNAVYK 1139
DB 1080 TGTQRQKNPEWELKLVNDKIVAHNVGSVPLSEYDNIQFSGKNMKDYSDSPKFTKLNNAVYK 1139
QY 1105 DTNNGFTQPLTNAFTASTYICLARENAEHGKINVTNKGFGQKS 1147
DB 1105 DTNNGFTQPLTNAFTASTYICLARENAEHGKINVTNKGFGQKS 1147
QY 1140 DIKSGFTQPLANAFST--GYISMARENAEHGKINANTKGGFGQKS 1181
DB 1140 DIKSGFTQPLANAFST--GYISMARENAEHGKINANTKGGFGQKS 1181

RESULT 10
US-08-053-614-2
; Sequence 2, Application US/08053614
; Patent No. 5403924
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: The taga gene and methods for detecting
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: NEEDLE & ROSENBERG, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/053,614
; FILING DATE: 19930426
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200,009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 859 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-053-614-2

Query Match 69.1%; Score 4056.5; DB 1; Length 859;
Best Local Similarity 92.7%; Pred. No. 1;le-249;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;

QY 1 MTNETIDQOPOTEAFAFPQOFINNLOVAFKVDNAVASYDDPOKPIVDKNDNRNRAQFEG 60
DB 1 MTNETIDQOPOTEAFAFPQOFINNLOVAFKVDNAVASYDDPOKPIVDKNDNRNRAQFEG 60
QY 61 ISOLREYSNKAIKNPTKKNQYFSDFINKSNDLINDNLIDVESSTKSKFQKGDQYRIF 120
DB 61 ISOLREYSNKAIKNPTKKNQYFSDFINKSNDLINDNLIDVESSTKSKFQKGDQYRIF 120
QY 121 TSWVSHQNDPSKINTSRINFMENIIQPIILDDKEAEFLKSAKQSFAGIIIGNQIRTDQ 180
DB 121 TSWVSHQNDPSKINTSRINFMENIIQPIILDDKEAEFLKSAKQSFAGIIIGNQIRTDQ 180
QY 181 KFMGVFDESILKQEAKEKNGEPTGGDWLDFISFIDKQSSDVKEAINOEPPHVPQDI 240
DB 181 KFMGVFDESILKQEAKEKNGEPTGGDWLDFISFIDKQSSDVKEAINOEPPHVPQDI 240
QY 241 ATTTTDDIOGLPPEARLDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNALS 300
DB 241 ATTTTDDIOGLPPEARLDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGCARHDWATVGYKQOQGNVATIIINVHMKNGSLV 360
DB 301 SVLMGSHNGIEPEKVSLLYGGNGGPGCARHDWATVGYKQOQGNVATIIINVHMKNGSLV 360
QY 361 IAGGEGKINNPSFYLYKEDQLTGSORALSOEIQNKIDFMEFLAQNNAKLDNLSEKEK 420
DB 361 IAGGEGKINNPSFYLYKEDQLTGSORALSOEIQNKIDFMEFLAQNNAKLDNLSEKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAQVSKDKTKHSAITFPGNGDLSYTLKDYGKKA 480
DB 421 FRTEIKDFQKDSKAYLDALGNDRIFAQVSKDKTKHSAITFPGNGDLSYTLKDYGKKA 480
QY 481 LDREKNVTLOGLSKHDGVNMFVDSYFNKYNASKNPNKGVGTNGVSHLGVFNKVAIFNL 540
DB 481 LDREKNVTLOGLSKHDGVNMFVDSYFNKYNASKNPNKGVGTNGVSHLGVFNKVAIFNL 540
QY 541 PDNLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
DB 541 PDNLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
QY 601 TGNDEVKKAQDKLEKSLKRRLHLEKEVEKLEKSGNKNKMEAKAQAANSQKDEIFALIN 660
DB 601 TGNDEVKKAQDKLEKSLKRRLHLEKEVEKLEKSGNKNKMEAKAQAANSQKDEIFALIN 660
QY 661 KEANDARAIAAQNKGKRELSDKLENVKNLDFDKSPDEFKNGKNKDFSKAEETLK 720
DB 661 KEANDARAIAAQNKGKRELSDKLENVKNLDFDKSPDEFKNGKNKDFSKAEETLK 720

Db 661 KEANRDARAIAAQNLIKIGIKRELSDKLENKDLKDFSKSFDGFKNGKNKDFSKAETLTK 720
QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDQENSIKDVIINQKI 780
QY 781 TKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKELQAQAKNESLNARKKSEIYQSV 840
Db 781 TKVDNLNOAVSVAKIACDFSGVEQALADLKNFSKELQAQAKNESFNV-CKSEIYQSV 839
QY 841 KNGVNGTLVGNLSQAEAT 859
Db 840 KNGVNGTLVGNLSGIEAT 858

RESULT 11
US-08-316-397B-2
; Sequence 2, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316.397B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 859 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-316-397B-2

Query Match 69.1%; Score 4056.5; DB 1; Length 859;
Best Local Similarity 92.7%; Pred. No. 1.1e-249;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;
QY 1 MTNETIDQOQTEAFNPQOFTNNLQVAFKVDNAVASYDDQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOQTEAFNPQOFTNNLQVAFKVDNAVASYDDQKPIVDKNDNRDNRQAFEG 60
QY 61 ISOLREYSNKAIKNPTKKNQVFSDFPINKSNDLKNKNDLIDVESSTKSPQKFGDQRYRIF 120
Db 61 ISOLREYSNKAIKNPTKKNQVFSDFPINKSNDLKNKNDLIDVESSTKSPQKFGDQRYRIF 120
QY 121 TSWVSHQNDPSKINTSRNFMENITQPPILDDEKAFLKSAKQSFAGIIGNQIRTDQ 180

Db 121 TSWVSHQNDPSKINTSRNFMENITQPPILDDEKAFLKSAKQSFAGIIGNQIRTDQ 180
QY 181 KPMGVFDESILKERQAEKNGEPTGGDWLDFLFSFIDKKOSSDVKKAINOEPVPHVQPD 240
Db 181 KPMGVFDESILKERQAEKNGEPTGGDWLDFLFSFIDKKOSSDVKKAINOEPVPHVQPD 240
QY 241 ATTTTIOGLPPARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFQQLLIHNHLS 300
Db 241 ATSTTHIOGLPPESRDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFQQLLIHNHLS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNTATVYKDOQGNNVATIIINHKNKSGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNTATVYKDOQGNNVATIIINHKNKSGLV 360
QY 361 IAGGEGKINNPSPYLYKEDOLTGSORALSOEEIQNKIDMEFLAONNAKLDNLSEKEK 420
Db 361 IAGGEGKINNPSPYLYKEDOLTGSORALSOEEIQNKIDMEFLAONNAKLDNLSEKEK 420
QY 421 FRTEIKDFOKDSKAYLDALGNDRIFAFVSKKDTKHSALITFEFGNGDLSYTLKDYGGKADKA 480
Db 421 FRNEIKDFOKDSKPYLDALGNDRIFAFVSKKDPKHSALITFEFGNGDLSYTLKVMGKKQIKA 480
QY 481 LDREKNVTLOGSLKHGDMVDFVSNFKYTNASKNPNKGVGTNGVSHLEVGNKVAIFNL 540
Db 481 LDREKNVTLOGNLKHDGMVDFVSNFKYTNASKNPNKGVGTNGVSHLEVGNKVAIFNL 540
QY 541 PDLNLAITSFVRRNLEDLTKTGLSPQANKLIKDFLSSNKELVGTILNFKNKAVADAKN 600
Db 541 PNLNLAITSFVRRNLEDLTKTGLSPQANKLIKDFLSSNKELVGTILNFKNKAVADAKN 600
QY 601 TGNVDEVKRAQKDLKSLKREHLEKVEKKLESKSGNKNKMEAKQAQANSOKDEIFALIN 660
Db 601 TGNVDEVKRAQKDLKSLKREHLEKVEKKLESKSGNKNKMEAKQAQANSOKDEIFALIN 660
QY 661 KEANRDARAIAAQNLIKIGIKRELSDKLENKDLKDFSKSFDGFKNGKNKDFSKAETLTK 720
Db 661 KEANRDARAIAAQNLIKIGIKRELSDKLENKDLKDFSKSFDGFKNGKNKDFSKAETLTK 720
QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDQENSIKDVIINQKI 780
QY 781 TKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKELQAQAKNESLNARKKSEIYQSV 840
Db 781 TKVDNLNOAVSVAKIACDFSGVEQALADLKNFSKELQAQAKNESFNV-CKSEIYQSV 839
QY 841 KNGVNGTLVGNLSQAEAT 859
Db 840 KNGVNGTLVGNLSGIEAT 858

RESULT 12
US-09-034-306-2
; Sequence 2, Application US/09034306
; Patent No. 5876943
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,306
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-034-306-2

Query Match 69.1%; Score 4056.5; DB 2; Length 859;
Best Local Similarity 92.7%; Pred. No. 1.1e-249;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;

Qy 1 MTNETIDQOQTEAAFNPOQFNNLQVAFKVDNAVASDPPQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOQTEAAFNPOQFNNLQVAFKVDNAVASDPPQKPIVDKNDNRDNRQAFEG 60
Qy 61 ISQLEEYSKAIKPTKKNQYFSDFINKSNDLKNKDLNIDVESSTKSFQKFGDQRYRIF 120
Db 61 ISQLEEYSKAIKPTKKNQYFSDFINKSNDLKNKDLNIDVESSTKSFQKFGDQRYRIF 120
Qy 121 TSWVSHQNDPSKINTRSINRMENIQQPIILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHQNDPSKINTRCIRNMEHTIQPIIPDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Qy 181 KFMGVFDESLEKQEAENKGEPTGDMWLDIFLSFIDFKKQSSDVKEAINEQPEVPHVQPD 240
Db 181 KFMGVFDESLEKQEAENKGEPTGDMWLDIFLSFIDFKKQSSDVKEAINEQPEVPHVQPD 240
Qy 241 ATTTTIDQGLPEARDLDERGNFSKFTLGDMWMLDVEGVADIDPNYKFNOLLIHNNALS 300
Db 241 ATSTTHIQGLPEARDLDERGNFSKFTLGDMWMLDVEGVADMDPNYKFNOLLIHNNALS 300
Qy 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNAATVGYKQQGNNVATTIINVHMKNGSLV 360
Db 301 SVLMGSHDGEIEPEKVSLLYAGNGGPGARHDWNAATVGYKQQGNNVATTIINVHMKNGSLV 360
Qy 361 IAGGEGKGNNSFYLYKEDQLTGSORALSQBEIQNKIDFMFEFLAONNAKLDNLSKEKEK 420
Db 361 IAGGEGKGNNSFYLYKEDQLTGSORALSQBEIQNKIDFMFEFLAONNAKLDNLSKEKEK 420
Qy 421 FRTEIKDFOKDSKAYLDALGNDRIAFVSKDQTKHSALITFEFGNGDLSYTLKDYGGKADKA 480
Db 421 FRNEIKDFOKDSKAYLDALGNDRIAFVSKDQTKHSALITFEFGNGDLSYTLKDYGGKADKA 480
Qy 481 LDREKNVTLOGSLKHGDMFYFDYSNFKYTNASKNPNKNGVGTNGVSHLEGFNKAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMFYFNYSNFKYTNASKNPNKNGVGTNGVSHLEAGFSKVAIFNL 540
Qy 541 PDLNNLAITSFVRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTYLNNKAVADAKN 600
Db 541 PDLNNLAITSFVRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTYLNNKAVADAKN 600
Qy 601 TGNVDEVKAKQDLKSLRKRREHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
Db 601 TGNVDEVKAKQDLKSLRKRREHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660

Qy 661 KEANDARAIYAQNKGKIKRELSKLENVNKNLKDOKSDFEFKNGKNKDFSKAETLK 720
Db 661 KEANDARAIYAQNKGKIKRELSKLENVNKNLKDOKSDFEFKNGKNKDFSKAETLK 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVVDVINOQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVVDVINOQKV 780
Qy 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLNKFSQEAQAQAQAKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLNKFSQEAQAQAQAKNESLNARKKSEIYQSV 840
Qy 841 KNGVNGTLVGNLSQAEAT 859
Db 840 KNGVNGTLVGNLSGIEAT 858

RESULT 13

US-09-259-437-2
Sequence 2, Application US/09259437
Patent No. 6153390
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOUS
APPLICANT: TUMMURU, MURALI K. R.
TITLE OF INVENTION: THE Taga GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/259,437
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-259-437-2

Query Match 69.1%; Score 4056.5; DB 4; Length 859;
Best Local Similarity 92.7%; Pred. No. 1.1e-249;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;

Qy 1 MTNETIDQOQTEAAFNPOQFNNLQVAFKVDNAVASDPPQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOQTEAAFNPOQFNNLQVAFKVDNAVASDPPQKPIVDKNDNRDNRQAFEG 60

```
QY 61 ISQREYSNKAIAKNTKKNQYFSDPINKSNDLNDLIDVESSTKSFQKFGDQYRIF 120
Db 61 ISQREYSNKAIAKNTKKNQYFSDPINKSNDLNDLIDVESSTKSFQKFGDQYRIF 120
QY 121 TSWVSHONDPSKINTSRINFMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHONDPSKINTSRINFMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
QY 181 KFMGVFDESLEKQERBAEKNGEPTGGDWLDIFLSFIDFKKQSSDVKEAINQEPVPHVQPD 240
Db 181 KFMGVFDESLEKQERBAEKNGEPTGGDWLDIFLSFIDFKKQSSDVKEAINQEPVPHVQPD 240
QY 241 ATTTTIDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 ATTTTIDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGPGCARHDNATVGYKDOQGNVATINVMHKNKSGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGPGCARHDNATVGYKDOQGNVATINVMHKNKSGLV 360
QY 361 IAGGEKGINNPSFYLYKEDQLTGSORALSQEBIQNKIDFMEFLAQNNAKLDLSSEKKEK 420
Db 361 IAGGEKGINNPSFYLYKEDQLTGSORALSQEBIQNKIDFMEFLAQNNAKLDLSSEKKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIAFVSKKDPKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 FRTEIKDFQKDSKAYLDALGNDRIAFVSKKDPKHSALITEFGNGDLSYTLKDYGKKADKA 480
QY 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKNPNKGVGTNGVSHLEVGFKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKNPNKGVGTNGVSHLEVGFKVAIFNL 540
QY 541 PDNNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNNKELVGTINFNKAVADAKN 600
Db 541 PDNNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNNKELVGTINFNKAVADAKN 600
QY 601 TGNVDEVKKAOKLEKSLRKRHELEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
Db 601 TGNVDEVKKAOKLEKSLRKRHELEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
QY 661 KEANDARAIAQAKLKGKRELSDKLENVANKLDFDKSDFDKNGKNKDFSKAEETLK 720
Db 661 KEANDARAIAQAKLKGKRELSDKLENVANKLDFDKSDFDKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKOLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVADVIINQV 780
Db 721 ALKGSVKOLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVADVIINQV 780
QY 781 TDVDNLQAVSVAKATGDFSRVEQALADLNFSKEQLAQQAQAKNESLNARKKSEIYQSV 840
Db 781 TDVDNLQAVSVAKATGDFSRVEQALADLNFSKEQLAQQAQAKNESLNARKKSEIYQSV 840
QY 841 KNGVNGTLVGNGLSQAEAT 859
Db 841 KNGVNGTLVGNGLSQAEAT 859
```

RESULT 14

PCT-US93-09782-2

Sequence 2, Application PC/TUS9309782

GENERAL INFORMATION:

APPLICANT: COVER, TIMOTHY L.

APPLICANT: BLASER, MARTIN J.

APPLICANT: TUMMURU, MURALI K. R.

TITLE OF INVENTION: THE tagA GENE AND METHODS FOR DETECTING

TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P.C.

STREET: 133 Carnegie Way, Suite 400

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

```
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09782
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200,009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-09782-2
```

Query Match

Best Local Similarity 69.1%; Score 4056.5; DB 5; Length 859;

Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;

```
QY 1 MTNETIDQPOPTAAAFNPQOFINNLOVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQPOPTAAAFNPQOFINNLOVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
QY 61 ISQREYSNKAIAKNTKKNQYFSDPINKSNDLNDLIDVESSTKSFQKFGDQYRIF 120
Db 61 ISQREYSNKAIAKNTKKNQYFSDPINKSNDLNDLIDVESSTKSFQKFGDQYRIF 120
QY 121 TSWVSHONDPSKINTSRINFMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHONDPSKINTSRINFMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
QY 181 KFMGVFDESLEKQERBAEKNGEPTGGDWLDIFLSFIDFKKQSSDVKEAINQEPVPHVQPD 240
Db 181 KFMGVFDESLEKQERBAEKNGEPTGGDWLDIFLSFIDFKKQSSDVKEAINQEPVPHVQPD 240
QY 241 ATTTTIDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 ATTTTIDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGPGCARHDNATVGYKDOQGNVATINVMHKNKSGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGPGCARHDNATVGYKDOQGNVATINVMHKNKSGLV 360
QY 361 IAGGEKGINNPSFYLYKEDQLTGSORALSQEBIQNKIDFMEFLAQNNAKLDLSSEKKEK 420
Db 361 IAGGEKGINNPSFYLYKEDQLTGSORALSQEBIQNKIDFMEFLAQNNAKLDLSSEKKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIAFVSKKDPKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 FRTEIKDFQKDSKAYLDALGNDRIAFVSKKDPKHSALITEFGNGDLSYTLKDYGKKADKA 480
QY 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKNPNKGVGTNGVSHLEVGFKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKNPNKGVGTNGVSHLEVGFKVAIFNL 540
QY 541 PDNNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNNKELVGTINFNKAVADAKN 600
Db 541 PDNNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNNKELVGTINFNKAVADAKN 600
QY 601 TGNVDEVKKAOKLEKSLRKRHELEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
Db 601 TGNVDEVKKAOKLEKSLRKRHELEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
```

```
Qy 661 KEARARATAYANLKGKIKRELSDKLENVKNLKDQDKSPDEKNGKNKDFSKAEETLK 720
|||||
Db 661 KEARARATAYANLKGKIKRELSDKLENINKDLKDFSKSPDFKNGKNKDFSKAEETLK 720
|||||
Qy 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSKVDVLIINQV 780
|||||
Db 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDQENSKVDVLIINQV 780
|||||
Qy 781 TDKNLNLQAVSVAKATGDFSRVQALADLKNFSKEQLAQQAQAKNESINARKKSEIYQSV 840
|||||
Db 781 TDKNLNLQAVSVAKIACDFSGVQALADLKNFSKEQLAQQAQAKNESFV-GKSEIYQSV 839
|||||
Qy 841 KNGVNGTLVGNLSQAEAT 859
|||||
Db 840 KNGVNGTLVGNLSGIEAT 858
|||||

RESULT 15
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCT
; CURRENT APPLICATION NUMBER: US/09/308,375
; EARLIER FILING DATE: 1999-05-14
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-308-375-2
```

```
Query Match 4.8%; Score 284; DB 4; Length 2285;
Best Local Similarity 19.3%; Pred. No. 1.3e-09;
Matches 259; Conservative 195; Mismatches 454; Indels 434; Gaps 63;

Qy 13 EAAF-NPQQFINNOVAFKLVNDNAV---ASYDPDQPKPIVDKNDRDNQAFEGISQLE- 66
|||||
Db 1010 EKAFDNKKDLQSLLETYSKSDSIDVFKMSDFDKAQNKKID-GDKSLSSVKSEVGDLCET 1068
|||||
Qy 67 --EYSNKAIKNPTKKNQYFSDFINKSNLDINKDLIDVESSTKSF---QKF-----G 113
|||||
Db 1069 LAEAGNEA-----EDFGKKLKEALDANSVDDIKAAIKEMSDAMQDSDVQDVNLG 1117
|||||
Qy 114 DORYRIFTSWYHQNDPSKINSTRINRPMENIOPPILDKKEAEFLKSAKQSFAGIITIG 173
|||||
Db 1118 D-----IFNN-----TKDOVAPLNDLIERMAGKSKISANEANTLIQKDELQALQAI 1165
|||||
Qy 174 NQIRTDQKFMGVFDESLKERQEAENKGEPTGGDWLIDFLSPI--FDDKQSSDVKEAINQOE 231
|||||
Db 1166 NGVVKINR-----DEVIKQRVK-----LDAYNDMVTYSNKLKMTFVNNAIK-- 1207
|||||
Qy 232 PVPHVQPDIAITTTDIOGLPPEARDLDERG-NFSKFTLGDMEMLDVEGVAD----- 282
|||||
Db 1208 -----TLNADTLRIDSL----KRLKERKLDWSEAELESDLEKVINNVADAKKELKKL 1256
|||||
Qy 283 ----IDP-----NYAFNOLLIIHNALLSVLMGSHNGI---EPEKVSLLYGGNGGPGCARHDW 331
|||||
Db 1257 EEKMLQPGGYSNQIEAMQSVKSALESYISASEATSTQEMNKQALVEAGT-----SLENW 1312
|||||
Qy 332 NATVGYDQOQGNVNVATIINVHMKNSGLVIAGGEKGINNPSFYLYKEDQLTGSQRALSQE 391
|||||
Db 1313 T-----DQO-----EKANETKYSMYVDK----- 1332
|||||
Qy 392 EIQNKIDMEFLAQNNAKLDNLSEKEKFKRTETKDFQKSKAYLDALGNDRIFVSK-- 449
|||||
```

Search completed: February 13, 2002, 14:02:16
Job time: 3405 sec

```
Db 1333 -----YKEALEKVNAEID-----KYNQOVNDYPKYSOKYRDAIKKEIKALQOQKK 1377
|||||
Qy 450 -----KDTKHSALITEFG-----NGDLSYTLKDYKKKADKALDREKNYVLOGS 492
|||||
Db 1378 LMQEQAKLLKQIKSGNITQYGVITSTTSSGGTSPSTGGSYSGK----- 1421
|||||
Qy 493 LKHDCVMFVDYSNFKYTNASKNPKNKGVGTNGCVSHLEVGFNKVA-----IFNL----- 540
|||||
Db 1422 -----YSSYINSAASKY-NYDPAIIAAVIOQESGFNAKARGVGAMGLQMLPAT 1470
|||||
Qy 541 -----PDLNLAITSEVRNLE-----DKLTTRKGLSP-QBA 570
|||||
Db 1471 AKSLGVNNAYDPQYNNMGTKYLAQOLEKFGGVEKALAAAYNAGPCNVIKYGIIPFKET 1530
|||||
Qy 571 NKLKIDFLSS-NKELVGKTLNPNKAVADAKNTGNYDEVKKAQKDLKSLRKRHEHLEKEVE 629
|||||
Db 1531 QNYVKKIMANYSKSLSSAT-----SSIASYTTNNSAPRVSSKYQGESGLRSPHPKGTDF 1586
|||||
Qy 630 -----KKLES-----KSGN-----KNKMEAKAQAANSOKDEI 655
|||||
Db 1587 AKAGTAIKSLQSGKQVQIAGYSKTAGNWWVIKODDGTVAKYMHMLNTPSYKAGQSVKAGOT 1646
|||||
Qy 656 FALINKEAN-----RDARAI-----AYAOLKGIKIKRELSDK 686
|||||
Db 1647 IGKVGSTGNTGNHLLHQLIEQBQKTIIDPEKYMGGTGTSTISDASQAEAEQOQIAQAKSDL 1706
|||||
Qy 687 L-----ENYNNKDKDFKSFDEFFK-----NGKNKDFSK- 714
|||||
Db 1707 LSLQGDISSVNDQIQOELQYELVQSKLDEFDKRIGDFDVRIAKDESANRYTSDSEFRKY 1766
|||||
Qy 715 AEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDL-----ENSV 771
|||||
Db 1767 TSDQKVAEQAQKIQOQKVNWIQKIKTNKALNSAQRAQLQE--ELKQAKLGLISVQDOV 1824
|||||
Qy 772 KDVIINQVTKVDVNLNOAV--SVAKATGDFSRVQALA-----DLKNFSKE-OLAQO 821
|||||
Db 1825 RE-LQKQLVQSKVDETLSKIEKSSKTOGKIKDVENKISMTEDEKVKYYSKQIKLIQO 1883
|||||
Qy 822 AQKNESLMARKSEIYQSVKNGVNGTLVGNLSQAEATTLTSKNFSQAEATTLTSKNFSQAEATTL 881
|||||
Db 1884 QKEAKKYIKOLEEQKKAAK-----GFPDIQEQITEEMQNKWKQKQDNLEL---Y 1931
|||||
Qy 882 NNNGLKNEPIYAKVYNNKKKAGQAASLEPIY-----AQVAKKVNAKIDRLNQIASGLGV 936
|||||
Db 1932 NTKKSIKD-----LYKSLADEWSIYKEMEKMRDIELEAHQKATQDLIDEI----- 1978
|||||
Qy 937 GOAAGFPLKRHDKVDLDSKV--GLSRNQLAOKI-DNLNQ-AVSEAKAGFFGNLEOT--- 989
|||||
Db 1979 -----DKTDEAKFOKELKERQDSIQKLTQDQINQYSLDDSEFGKSKVKELEQL 2027
|||||
Qy 990 -----IDKLKDSKHNPMNLWESAKKVPASLSAKLDNYATNSHIRINSNIKNGAINE 1042
|||||
Db 2028 QKEQLDLDDFLKDRSNKKEKALQOLEKDEESINKNYDNLVYND--RAFKLEDKIMNG 2085
|||||
Qy 1043 KATGMLTKQNPMLKLVNDKIVAHNVG-SVPLSEYDKIGFNQKNMKDYSDSPKSTFKLNN 1101
|||||
Db 2086 KITDIQAKQLN-EFSKFINTNM--ESIGKSIISNLLDK-----LKEASNA-----LNT 2129
|||||
Qy 1102 AVKDPTNSGFTQFLTNAFSTASY 1123
|||||
Db 2130 AVKGNTTGKK-----VSSSFASGGY 2148
|||||
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2002, 13:13:01 ; Search time 70.08 Seconds
(without alignments)
1246.750 Million cell updates/sec

Title: US-09-360-685A-5
Perfect score: 5867
Sequence: 1 MTNETIDQPPQTEAAFNPOQ.....RENAEHGIKNVNTKGFQKS 1147

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5823	99.3	1215	2 B48281	cytotoxin-associat
2	5395.5	92.0	1186	2 C64588	cag pathogenicity
3	5086	86.7	1167	2 B71924	cag island protein
4	349.5	6.0	1302	1 JC6009	surface-located me
5	346.5	5.9	2401	2 T28676	rhostry protein -
6	336.5	5.7	2269	2 T28677	rhostry protein -
7	331	5.6	1127	2 T28317	ORF MSV156 hypothe
8	322.5	5.5	2166	2 G70163	hypothetical prote
9	321	5.5	1365	2 T30822	lmp1 protein - Myc
10	317	5.4	1790	2 S67593	transport protein
11	307	5.2	1252	2 B42771	reticulocyte-bindin
12	302	5.1	2139	2 T18296	myosin heavy chain
13	301	5.1	1957	2 T38077	hypothetical colle
14	294.5	5.0	2523	2 T18477	hypothetical prote
15	293.5	5.0	1005	2 A64465	hypothetical prote
16	291	5.0	1269	2 F84730	probable myosin he
17	290.5	5.0	1738	2 T14867	interaptin - slime
18	290	4.9	1939	2 T18372	repeat organellar
19	290	4.9	1979	2 C71622	hypothetical prote
20	289	4.9	2663	1 S28261	centromere protein
21	286	4.9	1650	2 T18444	hypothetical prote
22	284	4.8	1875	2 S38173	myosin-like protei
23	284	4.8	2285	2 T12796	probable transglyc
24	279	4.8	1679	2 S48385	hypothetical prote
25	277.5	4.7	2288	2 T29999	hypothetical prote
26	276.5	4.7	2829	2 A42771	reticulocyte-bindin
27	276	4.7	1837	2 T41023	probable nuclear p
28	276	4.7	1928	2 S46773	myosin heavy chain
29	276	4.7	4688	2 F82885	hypothetical prote

30	275.5	4.7	2485	1 H71621	serine/threonine-s
31	272.5	4.6	2022	2 T43214	ovtl protein - nem
32	272	4.6	1938	1 MWKWI	myosin heavy chain
33	268	4.6	1208	2 T39068	coiled coil protei
34	267	4.6	1727	2 T50073	myosin-like coiled
35	267	4.6	1805	1 A64224	hypothetical prote
36	265.5	4.5	1413	2 T26467	hypothetical prote
37	265.5	4.5	1940	2 A59287	myosin heavy chain
38	265	4.5	1963	1 MWKW	myosin heavy chain
39	265	4.5	3394	2 T18501	hypothetical prote
40	264.5	4.5	1475	2 T33318	hypothetical prote
41	264.5	4.5	1957	2 A45627	myosin heavy chain
42	264	4.5	624	2 PC6003	surface membrane p
43	264	4.5	1313	2 F96673	hypothetical prote
44	263.5	4.5	3259	1 A56539	giantin - human
45	263	4.5	1164	2 T24806	hypothetical prote

ALIGNMENTS

RESULT 1
B48281
cytotoxin-associated gene A protein cagA - Helicobacter pylori (strains G39G and CCUG
N:Alternate names: immunodominant 128K antigen; immunodominant 135K antigen
C:Species: Helicobacter pylori
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: B48281; A48281
R:Covacci, A.; Censini, S.; Bugnoli, M.; Petracca, R.; Burroni, D.; Macchia, G.; Mass
Proc. Natl. Acad. Sci. U.S.A. 90, 5791-5795, 1993
A:Title: Molecular characterization of the 128-kDa immunodominant antigen of Helicoba
A:Reference number: A48281; MUID:93296225
A:Accession: B48281
A:Molecule type: DNA
A:Residues: 1-1215 <COV1>
A:Cross-references: GB:X70039; NID:g394912; PIDN:CAA49633.1; PID:g394913; GB:X70038;
A:Experimental source: strain G39G
A:Note: sequence extracted from NCBI backbone (NCBIN:133971, NCBIP:133973)
A:Accession: A48281
A:Molecule type: DNA
A:Residues: 1-957,1026-1215 <COV2>
A:Cross-references: GB:X70039; NID:g394912; PIDN:CAA49633.1; PID:g394913
A:Experimental source: strain CCUG 17874
A:Note: sequence extracted from NCBI backbone (NCBIN:133934, NCBIP:133968)
C:Comment: These sequences are from cytotoxin producing strains. It is similar to a s
C:Keywords: duplication; tandem repeat
F:958-1025/Region: 33-residue repeats

Query Match 99.3%; Score 5823; DB 2; Length 1215;
Best Local Similarity 94.4%; Pred. No. 3.9e-203;
Matches 1147; Conservative 0; Mismatches 0; Indels 68; Gaps 1;

Qy	1	MTNETIDQPPQTEAAFNPOQFINNLOVAFKVDNAVSYDDQKPIVDKNDKNDNRNQAFEG	60
Db	1	MTNETIDQPPQTEAAFNPOQFINNLOVAFKVDNAVSYDDQKPIVDKNDKNDNRNQAFEG	60
Qy	61	ISQLREYSNKAIKNPTKKNQYFSDFINKSNDLINKNLDIVESSTKSFQKFGDQYRIF	120
Db	61	ISQLREYSNKAIKNPTKKNQYFSDFINKSNDLINKNLDIVESSTKSFQKFGDQYRIF	120
Qy	121	TSWVSHQNDPSKINTRSTRNFNMENIIQPPILDDKAEFLKSAKQSFAGIIGNQIRTDQ	180
Db	121	TSWVSHQNDPSKINTRSTRNFNMENIIQPPILDDKAEFLKSAKQSFAGIIGNQIRTDQ	180
Qy	181	KFMGVFDESLEKROAEKNGEPTGGDWLDFLSFIFDKKQSSDVKEAINQEPVPHQVPI	240
Db	181	KFMGVFDESLEKROAEKNGEPTGGDWLDFLSFIFDKKQSSDVKEAINQEPVPHQVPI	240
Qy	241	ATTTTIDQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNALS	300
Db	241	ATTTTIDQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNALS	300

Qy 301 SVLMGSHNGIEPEKVSLLYGGNGGCGARHDMNATVGYKDDQGNVATINVHMKNGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGGCGARHDMNATVGYKDDQGNVATINVHMKNGSLV 360
Qy 361 IAGGEGKGNPVSFYLYKEDQLTGSORALSOEIQNKIDPMEFLAONNAKLDNLSEKKEK 420
Db 361 IAGGEGKGNPVSFYLYKEDQLTGSORALSOEIQNKIDPMEFLAONNAKLDNLSEKKEK 420
Qy 421 FRTEIKDFQKSKAYLDALGNDRIAFVSKKDPKHHSALITEFGNGDLSYTLKDYGGKADKA 480
Db 421 FRTEIKDFQKSKAYLDALGNDRIAFVSKKDPKHHSALITEFGNGDLSYTLKDYGGKADKA 480
Qy 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKPNKNGVGTNGVSHLEVGNKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKPNKNGVGTNGVSHLEVGNKVAIFNL 540
Qy 541 PDNLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKAVADAKN 600
Db 541 PDNLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKAVADAKN 600
Qy 601 TGNVDEVKKAQKDLSEKSLRKRHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
Db 601 TGNVDEVKKAQKDLSEKSLRKRHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
Qy 661 KEANDARAIAQAQNLKGKRELSOKLENVKNLKDFOKDFEPKNGKNKDFSKAEETLK 720
Db 661 KEANDARAIAQAQNLKGKRELSOKLENVKNLKDFOKDFEPKNGKNKDFSKAEETLK 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSYKDVIIINQK 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSYKDVIIINQK 780
Qy 781 TDKVDNLQAQVSVAKATGDFSRVEQALADLNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLQAQVSVAKATGDFSRVEQALADLNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
Qy 841 KNGVNGTLVGNLSQAEATTLTKNFSIDIKKELNAKLGNFNNNNNGLKNEPIYAKVKKK 900
Db 841 KNGVNGTLVGNLSQAEATTLTKNFSIDIKKELNAKLGNFNNNNNGLKNEPIYAKVKKK 900
Qy 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVGQAAGFPPLKRHKVDLDSKVG--- 957
Db 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVGQAAGFPPLKRHKVDLDSKVGSR 960
Qy 958 ----- 957
Db 961 VSPPIYATIDDLGPPFLKRHDKVDLDSKVGSRVSPPIYATIDDLGPPFLKRHKVD 1020
Qy 958 -----LSRNQELAQKIDNLQAQVSEAKAGFFGNLBTQIDKLKDKSTKHNPMNLWVESAKKV 1012
Db 1021 DLSKVLNRNQLAQKIDNLQAQVSEAKAGFFGNLBTQIDKLKDKSTKHNPMNLWVESAKKV 1080
Qy 1013 PASLSAKLDNATNSHIRINSIKNGAINKATGMLTQKNPEWLKLVNDKIVAHNVGVSVP 1072
Db 1081 PASLSAKLDNATNSHIRINSIKNGAINKATGMLTQKNPEWLKLVNDKIVAHNVGVSVP 1140
Qy 1073 LSEYDKIGFNQNMKDYSDSPKFSFKLNNVAKDTNSGFTQFLTNAFSTASYCYCLARENAE 1132
Db 1141 LSEYDKIGFNQNMKDYSDSPKFSFKLNNVAKDTNSGFTQFLTNAFSTASYCYCLARENAE 1200
Qy 1133 HGKKNVNTKGGFQKS 1147
Db 1201 HGKKNVNTKGGFQKS 1215

RESULT 2
C64588
cag pathogenicity island protein cag26 - Helicobacter pylori (strain 26695)
N:Alternate names: cytotoxin-associated gene A protein
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: C64588
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKee, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, A.; Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:9739467
A:Accession: C64588
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1186 <TOM>
A:Cross-references: GB:AE000569; GB:AE000511; NID:g2313663; PIDN:AAD07614.1; PID:g231

Query Match 92.0%; Score 5395.5; DB 2; Length 1186;
Best Local Similarity 89.9%; Pred. No. 1e-187;
Matches 1068; Conservative 26; Mismatches 51; Indels 43; Gaps 5;
Qy 1 MTNETIDQ-----QPTAEAFNPQQFINNLQVAFILKVDNAVASYDDQKPIVDKNDNRN 55
Db 1 MTNETIDQTRPTDQTSQTAFDPQQFINNLQVAFILKVDNAVASFDPQKPIVDKNDNRN 60
Qy 56 QAFEGISOLREYSNKAIKNPTKKNOYFSDFINKNDLINKDNLIDVESSTKSFQKFGDQ 115
Db 61 QAFDGISOLREYSNKAIKNPTKKNOYFSDIFDKSNDLINKDNLIDVESSTKSFQKFGDQ 120
Qy 116 RYRFTSVSHQNDPSKINTSRINFMENIIOPTDLDKEKAEFLKSAKQSFAGIIIGNQ 175
Db 121 RYQFTSVSHQNDPSKINTSRINFMENIIOPTDLDKEKAEFLKSAKQSFAGIIIGNQ 180
Qy 176 IRTDQKFGVGFDESUKERQAEKNGEPTGGDWLDIFLSPFIPDKQSSDVKEAINQEPVPH 235
Db 181 IRTDQKFGVGFDESUKERQAEKNGEPTGGDWLDIFLSPFIPDKQSSDVKEAINQEPVPH 240
Qy 236 VQPDATATTTDIOGLPPEAROLLDERGNFSKFTLDGMEMLDVEGVADIPNKFQNLIIH 295
Db 241 VQPDATATTTDIOGLPPEAROLLDERGNFSKFTLDGMEMLDVEGVADIPNKFQNLIIH 300
Qy 296 NNALSSVLGSHNGTEPEKVSLLYGGNGGPGARHDMNATVGYKDDQGNVATINVHMKN 355
Db 301 NNALSSVLGSHNGTEPEKVSLLYAGNGGFGDKHDMNATVGYKDDQGNVATINVHMKN 360
Qy 356 GSGLVIAAGEKGINNPSFYLYKEDQLTGSORALSOEIQNKIDPMEFLAONNAKLDNLSE 415
Db 361 GSGLVIAAGEKGINNPSFYLYKEDQLTGSORALSOEIQNKIDPMEFLAONNAKLDNLSE 420
Qy 416 KEKEKFRTEIKDFQKSKAYLDALGNDRIAFVSKKDPKHHSALITEFGNGDLSYTLKDYCK 475
Db 421 KEKEKFRTEIKDFQKSKAYLDALGNDRIAFVSKKDPKHHSALITEFGNGDLSYTLKDYCK 480
Qy 476 KADKALDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKPNKNGVGTNGVSHLEVGNK 535
Db 481 KADKALDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKPNKNGVGTNGVSHLEVGNK 540
Qy 536 AIFNLPDLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKAV 595
Db 541 AVFNLPDLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELAGKALNFKAV 600
Qy 596 ADAKNTGNYDEVKKAQKOLEKSLRKRHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEI 655
Db 601 AEAKSTGNYDEVKKAQKOLEKSLRKRHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEI 660
Qy 656 FALINKANRRARAIAYAQNLKGIKRELSDKLENVKNLKDFOKDFEPKNGKNKDFSKA 715
Db 661 FALINKANRRARAIAYAQNLKGIKRELSDKLENVKNLKDFOKDFEPKNGKNKDFSKA 720
Qy 716 EETLKALGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSYKDV 775
Db 721 EETLKALGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSYKDV 780
Qy 776 INQKVTDKVDNLNOAVSAKATGDFSRVEQALADLNFSKEQLAQQAQKNESLNARKKSE 835
Db 781 INQKVTDKVDNLNOAVSAKATGDFSRVEQALADLNFSKEQLAQQAQKNESLNARKKSE 840

QY 836 IYQSVKNGVNGTLVNGLSQAEATTLTKNFSDDIKKELNAKLGNNNNNNGLKN--EDIY 893
Db 841 LYQSVKNSVNTLVNGLSGIEATLAKANFSDIKKELNEKFKNF--NNNNGLKNSTPIY 899
QY 894 AKVKKKKAQAASLEPIYAOVAKVNAKIDRLNIOASGLGVGQAAGFPPLKRHDKVDL 953
Db 900 AKVKKKTKQVASEPIYTOVAKVNAKIDRLNIOASGLGVGQAAGFPPLKRHDKVDL 959
QY 954 SKVGL-----SRNQLAQAQIDNLNQAVSEAK 979
Db 960 SKVGLSASPEPIYATIDDLGGPFPLKRHDKVDLTKSVGRSRNQLAQAQIDNLNQAVSEAK 1019
QY 960 AGFFGNLEOTIDKLKDSKTHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSIKNGA 1039
Db 1020 AGFFGNLEOTIDKLKDSKTHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSIQNGA 1079
QY 1040 INEKATGMLTKQNPWLKLVNDKIVAHNVGVSPLSEYDKIGFNQKNMKDYSDSFKSTKL 1099
Db 1080 INEKATGMLTKQNPWLKLVNDKIVAHNVGVSPLSEYDKIGFNQKNMKDYSDSFKSTKL 1139
QY 1100 NNAVKDTSNGTQFLTNAFSTASYCLARENAEHGKKNVNTKGGFQKS 1147
Db 1140 NNAVKDTSNGTFLANAFST--GYCLARENAEHGKKNVNTKGGFQKS 1186
RESULT 3
B71924
cag island protein, cytotoxicity associated immunodominant antigen - Helicobacter pylori
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: B71924
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path-
A:Reference number: A71800; MUID:99120557
A:Accession: B71924
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1177 <ARN>
A:Cross-references: GB:A6001483; GB:A601439; NID:94155034; PIDN:AAD06073.1; PID:9415503
A:Experimental source: strain J99
C:Genetics:
A:Gene: cagA

Query Match 86.7%; Score 5086; DB 2; Length 1167;
Best Local Similarity 84.1%; Pred. No. 1.5e-176;
Matches 996; Conservative 60; Mismatches 73; Indels 56; Gaps 4;
QY 1 MTNETIDQOPQTEAAFPNQFNNLQVAFKVDNNAVSYDPDQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNEAINOQOPQTEAAFPNQFNNLQVAFKVDNNAVSYDPDQKPIVDKNDNRDNRQAFEG 60
QY 61 ISQLEEVSNRAIKNPTKKNQYDFINKNNDLINKNLDIVESSTKSFQKPGDQRYRIF 120
Db 61 ISQLEEFANKAIKNPTKKNQYDFINKNNDLINKNLDIVESSTKSFQKPGDQRYRIF 120
QY 121 TSWSHQNDPSKINTRSIRNFMENIIQPIIDDKKAEFLKSAKQSFAGIIGNQIRTDQ 180
Db 121 MNWVSHQNDPSKINTOKIRGMENIIQPIIDDKKAEFLKSAKQAFAGIIGNQIRSDQ 180
QY 181 KPMGVFDESLEKQBAEKNGE-----PTGCDWLDIFLSFTFDKKQSSDVKEAINQEPVPHV 236
Db 181 KPMGVFDESLEKQBAEKNGEPNGDPTGCDWLDIFLSFVNKKQSSDLKETLNQEPVPHV 240
QY 237 QPDIAATTTDIOGLPEARDLLDERGNSKFTPLGDMEMLDVEGVADIDPNYKFNQLLIHN 296
Db 241 QPDVATTTDIOGLPEARDLLDERGNSKFTPLGDMEMLDVEGVADIDPNYKFNQLLIHN 300
QY 297 NALSSVLMGSHNGIEPEKVSLLYGGNGGPGARHDNATVGYKQDQGNVATLINVHMKN 356
|||||

Db 301 NALSSVLMGSHNGIEPEKVSLLYGGNGGPGARHDNATVGYKQDQGNVATLINVHMKN 360
QY 357 SGLVIAGCEKGINNPSFYLYKEDQLTQSORALSQBEIQNKIDFMEFLAONNAKLDNLSEK 416
Db 361 SGLVIAGCEKGINNPSFYLYKEDQLTQSORALSQBEIQNKIDFMEFLAONNAKLDNLSEK 420
QY 417 EKEKFPTEIKDFQKDSKAYLDALGNDRIAFVSKKDTKHSALITEFGNGDLSVTLKDYKK 476
Db 421 EKEKFPQNEIDFQKDSKAYLDALGNDRIAFVSKKDTKHSALITEFGNGDLSVTLKDYKK 480
QY 477 ADKALDRDKNVTTLOQSLKHGDMFVDYSNFKYTNASKPNKGVGTNGVSHLEVGFNKA 536
Db 481 ADKALDRDKNVTTLOQSLKHGDMFVDYSNFKYTNASKPNKGVGTNGVSHLEVGFNKA 540
QY 537 IFNPLDLNLAITTSVRRNLEDKLTGSLSQEANKLIKDFLSSNNKELVGKTLNFKAVA 596
Db 541 VFNPLNLAITTSVRRNLEDKLTGSLSQEANKLIKDFLSSNNKELVGKTLNFKAVA 600
QY 597 DAKNTGNYDEVKAKQDLKSLKREHLEKVEKLEKSGKNGKNKMEAKAQAANSODEIF 656
Db 601 EAKNTGNYDEVKAKQDLKSLKREHLEKVEKLEKSGKNGKNKMEAKAQAANSODEIF 660
QY 657 ALINKENRDAIRAIYAQNKLKIGIKRELSDKLENVKNLKDQKSFDEFKNGKNKDFSKAE 716
Db 661 ALINKENRDAIRAIYAQNKLKIGIKRELSDKLENVKNLKDQKSFDEFKNGKNKDFSKAE 720
QY 717 ETLKALKGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAQSDLENSVDVII 776
Db 721 ETLKALKGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAQSDLENSIKDVII 780
QY 777 NOKVTDKVDNLNQAVSVAKATGDFSRVQALADLNKFSKEQLAQQAQKNESLNARKKSEI 836
Db 781 NOKVTDKVDNLNQAVSVAKATGDFSRVQALADLNKFSKEQLAQQAQKNEDFTGKNSAL 840
QY 837 YQSVKNGVNGTLVNGLSQAEATTLTKNFSDDIKKELNAKLGNNNNNNGLKN--EDIY 896
Db 841 YQSVKNGVNGTLVNGLSQAEATTLTKNFSDDIKKELNAKLGNNNNNNGLKN--EDIY 894
QY 897 NKKAGAASLEPIYAOVAKVNAKIDRLNIOASGLGVGQAAG----- 941
Db 895 -----TEPIYTOVAKVNAKIDRLNIOASGLGVGQAAG----- 943
QY 942 -----FPLKRHDKVDLTKSVGRSRNQLAQAQIDNLNQAVSEAK 1003
Db 944 GLSANHEPIYATIDDLGGPFPLKRHDKVDLTKSVGRSRNQLAQAQIDNLNQAVSEAK 1003
QY 983 FGNLEOTIDKLKDSKTHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSIKNGAINE 1042
Db 1004 FGNLEOTIDKLKDSKTHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSIKNGAINE 1063
QY 1043 KATGMLTKQNPWLKLVNDKIVAHNVGVSPLSEYDKIGFNQKNMKDYSDSFKSTKLNA 1102
Db 1064 KATGMLTKQNPWLKLVNDKIVAHNVGVSPLSEYDKIGFNQKNMKDYSDSFKSTKLNA 1123
QY 1103 VKDTSNGTQFLTNAFSTASYCLARENAEHGKKNVNTKGGFQKS 1147
Db 1124 VKDTSNGTQFLTNAFSTASYCLARENAEHGKKNVNTKGGFQKS 1167
RESULT 4
JC6009
Surface-located membrane protein lmp3 precursor - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: JC6009
R:Ladefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.
J. Bacteriol. 178, 2775-2784, 1996
A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene sys-
A:Reference number: JC6009; MUID:96213016
A:Accession: JC6009
A:Molecule type: DNA
A:Residues: 1-1302 <LAD>
A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CAA64858.1; PID:g1197336

C:Genetics:

A:Gene: Imp3
 C:Genetic code: SGC3
 C:Superfamily: surface-located membrane protein Imp3; tetratricopeptide repeat homology
 C:Keywords: duplication; membrane protein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1302/Product: surface-located membrane protein Imp3 #status predicted <MAT>
 F:957-992/Domain: tetratricopeptide repeat homology <TTL>
 F:993-1026/Domain: tetratricopeptide repeat homology <TT2>
 F:1089-1120/Domain: tetratricopeptide repeat homology <TT3>
 F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match 6.0%; Score 349.5; DB 1; Length 1302;
 Best Local Similarity 19.8%; Pred. No. 1.4e-05;
 Matches 243; Conservative 230; Mismatches 478; Indels 277; Gaps 56;

QY 31 KYDNAVASDPDQKPIVDKNDNRDQAFEGISQLREESYKAIKPTKKNQYFDFINKS 90
 DB 58 KANDEFARODKFNSTAFKNHSNTSKLIDEI-----EIFSKTILELQDKORLEELNKL 113
 QY 91 NDL-INKNLDIVESTSKSFQKQRYRIFTSWYSHQNDPSKINTSRNFM-ENIIP 148
 DB 114 NKLRDLQNL-----NSNDGQNVDSDAKKALENQIDD 148
 QY 149 PILDDKEKA--EFLKSAKOSFAGIIGNOIRTDQKFMG-VFDESILKQEAKEKGEPTGG 205
 DB 149 SLPIDKIRKTNLENLAKKELL-----NKNAERELQSKIFNE---KKQELKRLVLDLED 200
 QY 206 DWLIDFLSIFDPKQSSDVKEAINQ-----EPVPH--VQPDATTT-----TD 246
 DB 201 KEVD-----FTKEQKQVETINETSIEDIKKIIIEVKATSSLTSKILMTKQOELQE 254
 QY 247 IQGLPEARDLDERGNSKFTLGDMEMLD-VEGVADIDPNYKFNOLLHNNALSSVLGM 305
 DB 255 FENIRKQLQDFINTKLNDKQYSIKQKALDINSUNGINKNSTIKEIKAGQNAL--IKAK 312
 QY 306 SINGTEPKVLLYGGNGPGARHDMNATVGYKDOOGNVATIIIVHMKSGSLVGIAGGE 365
 DB 313 EPAGLEKEK--LDGONIKDTLKETINNAKEFKLLIDNDQKIVD--LKSNDLNEISKAE 367
 QY 366 KGINNPVLYKEDOLTCORALSQESQETONKIDFNEFLAONNAKLDNLSEK--EKFT 423
 DB 368 QSLSKDKESMESANDLLNT-KLIEYKEILNK-----FNOEKEAFNELEQTRKNIENFLT 421
 QY 424 E-----IKDF--QKDSKAVLDALGNDRIAFVSKD--TKHSALITEFNGDLSY 468
 DB 422 DEVKNNPNYATLVKDLTNKADAKSVTNSN-----KSDIIAANEALIOALADANKA- 473
 QY 469 TLKDYCKKADKALDRKVV-----TLOGSLKHGQVMFV-----DYSNFKYTNASK--NPNK 517
 DB 474 --KQVDEANKSIKQLNALIDKANTLLPQLNDNDSEIVKAKESLNAEITNANKAVNQND 531
 QY 518 GVGVTNGVSHLEVGNPKVAIFNPLDNLNLAITSE-----VRNLEDKLTTKGLSPOEANK 572
 DB 532 NASWQSAKSSLDKVKTKIQ-NOLTFNFKDKDAKFELEQTRKIDNLFNTDDVKNKNPNYAT 590
 QY 573 LIKDFLSSNKELVGTILNFNKAADAKTNGYDEVKKAQKDLKSLRKRHLE---KEVE 629
 DB 591 LVKDLTNKADKKSVYKSSNKSEIIAAN---DELKQA--LDRAKAKAQDIDEANKSIK 643
 QY 630 KKLKESKNNKMEKAKAQSOKDEIFALINK---EANRDARATAYAOVL-----KGIG 680
 DB 644 EQLSDSITNANQLNKL-VDSKDK-----IQAKTELSEIQSASQELNLPNNTSMQSAK 697
 QY 681 RELSKLENVKNLKDQKDFDEFKNGKNKDFSKAEETLKALKGSVKDLGINPENW---IS 737
 DB 698 ESLDAKVTEITKLETFNKD-----KDVFKFELEKTRKIDIDFINTKNTNPNYSTLIS 750
 QY 738 KVENLNAALNEFKNGKNDKSKVTQAKSDLENSVKDVIINGKVTDKV-----DNLNQA 792
 DB 751 ELTSKRSKNSVTNSNK--SDIETANTELQALAKAKANTDKAQNLAQKSTKEQLNNSIS 808

QY 793 -----VAKATGDFSRVEQALADLKNFSDQLAQQAQKNESLNARKKSEIYQSVKNGVNGT 847
 DB 809 SANTLLAKLTKDNTIQQA-----KTELEKEVQKANAQVASNNTASMQSAKSLD-- 858
 QY 848 LVNGLSQAQEAATLSKFNFSDFIKELNAKLCNFNNNNGLKNEPIYAKVYKKNKKAQAASL 907
 DB 859 -----AKVTEITTKLETFNKDKDKVFKFELEQTR-----KDIDEFINTNK----- 897
 QY 908 EPIYAQVAKVNAKIDRLNQIASLGVVVG-QAAGFPPLKR-----HDKVDDLSK--- 955
 DB 898 TNPDYSTLISELTSKRSKNSITNSNKSIDIETANTELQALAKAKANTDKQADNLAARSTK 957
 QY 956 -----VGLSRNQELAQKIDNLNOAVSEAKAGFCNLEQITIDKLKSTKHNPMMLWVESAK- 1010
 DB 958 EQLNKSISANTLLAKLTKDNTIQQAKT-----ELEKEVQKANAQVASN-NTASMQSAKS 1012
 QY 1011 -----KVPASLSAKLDNYATNSHIR-----INSNIKNGAINEKATGMUTQKN 1052
 DB 1013 SLDAKV-TEITTKLETFNKDKDKVFKFELEQTRKIDIDEFINTN-----KTN 1056
 QY 1053 PEWLKLVNDKIVAHNVGVSPLSEYDKIGFNOKNMKDYSDSPKFTKLNNAVKTNSGFTQ 1112
 DB 1057 PNYSTLISE-----LTSKRSKNSITNSNKSIDIETANTELQALAKAKANTDKAQ 1105
 QY 1113 FLTNASTASYCYCLARENAEHGKKNVT 1140
 DB 1106 ADNLARST-----KEQLNKSISANT 1126

RESULT 5

rhoptry protein - Plasmodium yoelii (fragment)

C:Species: Plasmodium yoelii

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000

C:Accession: T28676; A45521

R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.

Mol. Biochem. Parasitol. 76, 329-332, 1996

A:Title: Comparison of two members of a multigene family coding for high-molecular ma

A:Reference number: 220507; MUID:97077455

A:Accession: T28676

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2401 <SIN>

A:Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1

R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42, 241-246, 1990

A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple

A:Reference number: A45521; MUID:91101660

A:Accession: A45521

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 2260-2401 <KEE>

A:Cross-references: GB:M34281

Query Match 5.9%; Score 346.5; DB 2; Length 2401;
 Best Local Similarity 20.0%; Pred. No. 3.7e-05;
 Matches 303; Conservative 224; Mismatches 462; Indels 525; Gaps 77;

QY 21 FINNL-QVAFUKVONAVASYDDQKPIVDKNDNRDQAFEGISQLREESYKAIKNTPK- 78

DB 54 FFNSLGLKLLINKVDS-----DGNIIIEENDIDN-----FDLSKPKNNEKL 92

QY 79 -KNQVFSDFINKNDLKNLDIVESSTKSFQKQRYRIFTSWYSHQNDPSKINTSR 137

DB 93 LEGELNGVFENKNDYKNNKL-----DQFKDTMKKILLIILQIYMNEFFGLNDM 142

QY 138 IRNFMENTIQQPIILDDKEAEFLKSAKOSFAGIIGNOIRTDQKFMGVFDESLEK----- 192

DB 143 TKLKNEGISQKRVINNOIKQKFDKST-----YDEKKEG-FESSLELAKNWE 187

QY 193 RQEAEE-----KNGEPTGGDWLIDFL-----SFIQKQSSDVKEAINQ-----E 231

Db 188 KKKLEIITELKKKNETV---OLDIKIRELIKQIKDIEEQIVNDLKLELNKKIKEITE 244
QY 232 PPHV-----OPDIATTTDIOGL---PPEARDLLDERGN-----FSKFTLGDM 273
Db 245 KIEYIKKAVDLKKEKONVYIDELAKEPPYQITKYEKKNEITYNTIKSDFDKIYVGDIE 304
QY 274 ML-----DVEGV-----ADIDPNYKEN 290
Db 305 QLYNEMFVQSSNIEHINKTEITLTKTIDNVYNNIONMETETVKSHLNKTETNNKLS 364
QY 291 QLLIH-----NNALSSVL-----MGSHNGIE---PEKVSIL-LYGGNGGPGARHD 330
Db 365 ETILDIKIYIGEITNELNKTLEDKNEKEGLSNKIDYAKENVQLNVYKSNILELKKH- 423
QY 331 WNATVGYKDOQGNVATIIV-----HMKSGSLVIAAGGEGKGINNPSF 373
Db 424 -----YNDQ-----INIDNIKEAKONYDOFKHEMK-----TIPPNEMKYQKPSI 464
QY 374 YL-----YKE-----DOLTGS-QRALSOEI---ONKI--- 397
Db 465 EIKIMKDEFLSKVKNYNDPKYKKEVSEHNKFTLTWKITEVSDIEIKKYENKFNDS 524
QY 398 -----DFMEFLAQNNAKLDNLSEKE---KEKFRTEIKDPQ 429
Db 525 KSLINETKKSIEEYONINTLKKVDDYIKVCLNTNELITNCHNKQTTLDKLNQNLKTIK 584
QY 430 KDS---KAYLDALGNDRIAFVSKDKTKHSALIT---EFGNGD-LSYTLKDYKGGKADKALD 482
Db 585 ETNSIDKIYTDKFENILTDKKTELETFTGLSLNNHESNNKELLTYF---YDLKANLGN 641
QY 483 REKNYTLQGLSHDGMFVDYSNFKYNASKPNKGVGTNGVSHLEVGFKVAIFNLP- 541
Db 642 KENMLYKOPNEKEKAVEDI-----KKKNVDINKIVSNIEITIT- YTSYININE 687
QY 542 -----DLNLAITSFVRN-----LEBKL-----TTKGLSPOEANKLIK 576
Db 688 DTENEIGKSIELNLTVKLEKRVANTYNLNEIKEKLDYDFQDFGKEKNKYPOENKIKND 747
QY 577 FLSSKNELVGKTLNFKAVADAKN-----TGNVDEVKKA-----QKDLKSLR--- 619
Db 748 IDTLNOKIDKSIETLTETKIKSENHIDEIKQIDKLKVPNKTMTFNEDEPKETEKENIV 807
QY 620 ---KREHLEKEVEK-----KLES-----KSGNKKMEAKAQA---SQDEIFALLN 660
Db 808 EKIDKKNIYKIDKLLNLEISINDKTSLEKLNINLSYKSLGNLFLOQIDE-----EK 863
QY 661 KEANRDARAI-AYAQNLIKIKRELSDKLE-----NVNKNLKDFDKGDFDEPK--- 705
Db 864 KRAEHTIKAMEAYIDDLNLIK-K-SOEIEKEMNINMDIKMDIHENKALNISHDDYKIIYH 922
QY 706 -NGKN-----KDFS-----KAEETLKALGSKVKGOLGINPEWISKYE 740
Db 923 TTSKNHEEKISDIRKNSLIIQDFSESVINDIKKEKNVLESQNNNTDIN-QYLSKIE 981
QY 741 NLN---AALNEFGNKNKFSKYTOAKSDLENSVKOVIINQKVTDKV-----DNLNQAV 791
Db 982 NIYNILKLNKIK---KIIDKYKEYTDETEKNKKKINAELSSEKIIITQKENSLSKECQ 1037
QY 792 SVAKATGDFSRYEQAALADLKNF-----SKEOL-----AQQAQKNESLN-----ARKK 833
Db 1038 SKIKSTIDNYYSECIKNITNLKTYIVNEKNINIYFKNAEEVNVQSLNFNNIEMADTK 1097
QY 834 SEIYOSVK---NGVNGTLVGNLSQAPEATTLKSNFSDIKKELNAKLGNNFNNNNNKLNEP 891
Db 1098 SQYILNIKKNGTNTDYN-----IKELHEKKKSNNVYKDEACKNTQEIKNKKE 1146
QY 892 IYAKVN-----KKKAGQAASLEPIYAQV-----AKVNV--- 920
Db 1147 LFEKVEQOEVTVLLNKYIAVELNKKFKDKTKYSEQIILKEIKDAHNTFTSQADKSEKKMNEI 1206
QY 921 -----AKIDRLNQAISGLGVGQAAGFPLK-RHDKYDDL---SKVGLSRNQELA 965
Db 1207 KNEQIRIEDEAVAKNNSKAILDI-----QLSVEPPKIRFLKTKIDRTKSDOCLKETKDIE 1262

QY 966 QKIDNUNQAVSEAKAGFFCNLEQFIDKLDKSTKHNPMLWVESAKKVPASLSAKLDNYAT 1025
Db 1263 TKISNLSIDTOETKLTENKNILNLTLEKLSLKNOKN---IEDOKKELDEVNSKIKNIES 1320
QY 1026 NSHIRINSKINCAINEKATGMLTOKNPWLKLVDN-----KIVAHNNGSV--PLSEYD 1077
Db 1321 N---NQHKKNVEIG-----IVEKINEIAKANKQOISTQKLIITPTKNLISPFPKAND 1370
QY 1078 KIGFN-QKNMKDYSDSFKEFTKLNNNAVKDTNSOFTLNAFSTASYICLARENAEH--- 1133
Db 1371 LEGIDTNKNLG-----KYNTENMNIYEFIKSY-DLITHYLETVSKEPITYBOIKNKRI 1423
QY 1134 -----GIKNVN 1139
Db 1424 TAQNELLTNIKVN 1437
RESULT 6
T28677
rhoptry protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28677; C45521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A:Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii
A:Reference number: Z20508; MUID:95021522
A:Accession: T28677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KE>
A:Cross-references: EMBL:L27838; NID:g457145; PID:g457146; PIDN:AAA21304.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple
A:Reference number: A45521; MUID:91101660
A:Accession: C45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

Query Match 5.7%; Score 336.5; DB 2; Length 2269;
Best Local Similarity 19.5%; Pred. No. 8e-05;
Matches 243; Conservative 232; Mismatches 479; Indels 291; Gaps 59;

QY 22 INNIQVAFILKVDNAVASYDDQKPIVDKNDNRQAFEGISQLRREYSKAIKNPKKNQ 81
Db 234 LNNIETN-NKLPNTLEI---KRYIYDEISKELNKLMLDFKKNKEKLSNK-ISDYDKRE 288
QY 82 YFSDFIKS-----NDLINKDLIDVESSTKSFQKFGDORYRIFTSWYSHONDPSKI- 133
Db 289 QLSEYKSKMLEIRNHYNSTQNTVDNTE-EEAKONYDKSNEH-----MTTPTNDEDEISKII 343
QY 134 -NTRSRNFEMENIQQPIILDDKEAEFLKSQSFAGIIIGNOIRTDQKFMGVFDESKE 192
Db 344 SEVKTMKDELKSVNTYIDFNKKYKETVNSEHSQFTEL-----TDKIAKAYSVDRELKK 396
QY 193 ROEAENGEPGTDGWLIDFLSFIDFKQSSDVKEAINQEPVHVQPDIAATTDDIOGLPP 252
Db 397 CEQ-----SFNDKNSLINETKNSEKE-----YONINT---LKKVDE 430
QY 253 EARDLLDERGNFSKFT---LGDMEMLDVEGVAD---IDPNY--KFNOLLIH---NN 297
Db 431 YKVCSTKESITKFSKOTILKMLNQNIKTVKETSIDKSTIERPEQLITGQTKLEN 490
QY 298 ALSSVLGMSHGNGTEPKGVSLLYGGNGGPGARHDNATVGY-----KDOQGNV 345
Db 491 KTEFSLNNHEANNNELIKYF-----SDKANLGINEENMLYNQFTEKEKTFNDI 540
QY 346 ATINVHMKNGSLVIAGGEGKGINNPSFYLYKEDQITGSORALSQBEIQNKIDFMFLAQ 405

Db 541 KE-KNIHINE---EISKIEIKI-HASIYINISEE--TEREIGINIESLNTKV--FEKVKE 590
Qy 406 NNAKLDNLSEKEKEKFRTEIKDFODKSK--AYLDALG--NDRIAFAVSKDKTKHSALETIFG 462
Db 591 NVTNLNKEKEKLK---HYDFSFGEGNTKYTDKIKKINDIMAVSQQIDQHINGLDDIQ 647
Qy 463 NGDLSY--LKDYKKADKALDRE-KNVTLOGSLKHGVMFYDYSNFKYTNAKSNPNKGV 519
Db 648 KKSSEYVSEMKEQINKLEKVSNTESNDNVEGKKKKQIIVTKID--KKKNIYEINKLL 705
Qy 520 GVTNGVSHLEVGFNKVAIFNPLDNLNLAITSFVRNLEDKLTGKLSPOEANKLIDFSL 579
Db 706 SEISKIEKNTSLEKVKDINLSYQNL-----GNLFLEQID 741
Qy 580 SNKELVGKTLNFKAVADAKNTGNVDEYKAKODLEKSLRKEHLEKEVE-----K 630
Db 742 EKKKAENTIKWEAYID-----DLNKKKQSEIETEMDIKMDINKEMALKISHDDK 796
Qy 631 KLESKSGN-----KNKMEAKAQAANSOKOEIFALINKE 662
Db 797 KCHDKSKNHNKENSIDYDKSSKIIQDFSRESINDIKNKLQ--KNVSESQNH-----NSD 849
Qy 663 ANRDARATAYAOANLKGIRKRELSOKLENNVKNLKDQSPDFKNGKNDKDFSKAEETLKL 722
Db 850 INOCLNEVANIYNILK-----NKIKTIIDKYKEYTSEIEKNKKNINDELNNSKVIK 904
Qy 723 KG--SVKDL--GINPEWISK-----VENLNAALNEFKNGKNDKFSKVTOAKSDLENSVKD 773
Db 905 EGDLSKECRKINSTLDDKDIKINLVK---KNILNEE--TNITNHFNAEYENKI 960
Qy 774 VIINQKVDKVDNLNOAVSVAKATGDFSVEQALADLKNFSKEQLAQQAQKNESLNA--RK 832
Db 961 VLSNFENNTEADNKSQYILEIKKNGTNDHDYNIKELSKDKSNGYKTEADQONKKAIOK 1020
Qy 833 KSEIYQSVKNGVNTLVGNLSQAEATTLKSNFSDIKELNAKLGNFNNNNN-----884
Db 1021 NKELFEQYKEEV--TVL---LNKYAVELKNKFDKTKNDKSDKQIIEIKDAHNYCTLES 1075
Qy 885 -----NLKNEPIYAK---VNKKKAGQA--ASLEEPIYQVAKKVNKADRLNNOIASGLCV 935
Db 1076 SEKKNEIKNEKIHIEDEVANNKSNKAITSVSV-----EPFKTKIINEIET----1126
Qy 936 VQQAAGFLPKRHDKVDLDSKVLGSLNQELAQKIDMLNOAVSEAKAGFGNLEQTIDKLKD 995
Db 1127 -----KSDCC-----LKETNDLEKQISNLSIDTQETKLTENGKQKLTLEELLE 1169
Qy 996 STKHPMNLWVESAKKVPASLASKLNDYATNSHIRNSIKN-----GAINKATGMLT 1049
Db 1170 SUKKOKKN--IEDQKKELDVNSKIKNIET-----VNOHKKNYEIGIVEKINEIAK---T 1220
Qy 1050 QKNP-EWLKLVNDKIVAHNVGVSPLSEYDKIGFNQKNMK--DYSDSF-KFSTKLNNAVKD 1105
Db 1221 NKNOIESTKELIKPTIOHIISS-----FNANDEGSDENLGDKNKYTEMGNIEE 1270
Qy 1106 TNSGTQELTNFAFSTASYICLARENAEH-----GIKNVN 1139
Db 1271 FTKSY-NLITNVLETVSRESITYIQNKRIDTQKELLKNIENVN 1314

RESULT 7
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28317
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: #20484; MUID:99102612
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-1127 <AFO>
A:Cross-References: EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:g4049717
C:Genetics:
A:Note: MSV156

Query Match 5.6%; Score 331; DB 2; Length 1127;
Best Local Similarity 20.7%; Pred. No. 5.4e-05;
Matches 249; Conservative 177; Mismatches 440; Indels 338; Gaps 53;

Qy 53 DNRQAFEGISQUREEYKNAIKNPTKKYQFSDFINKSNLDINKONLIDVESTSKSFQRP 112
Db 25 NNVSLDIINSLEYILNN--IK-----FSDKI--TNEIIKNYKNI-VEIFYMHQF 71
Qy 113 GQORYRIFTSW-VSHONDPSKI---NTRSIRNFMEIIQPPIL-----DQKEAEFLK 161
Db 72 KINDYNILOIYIEYNEINEIKCIENKPKCKNPLNYITYKKLYIYDLDYEEKKOKELVI 131
Qy 162 SAKQSFAGIIIG-----NQIRTDQKFMGVFDESLKEROEAEKNGEPTGDMLDIFLSPI 215
Db 132 NIEQKNAVDKINDIKNNVNNIHSNDETIITGKETL-----IDILNKL 174
Qy 216 FDKKSSDVKEAINQ-----EPVPHVQDPIATTTDIOGLPPEA--RDLDERG 262
Db 175 --KLVSDEKQLEIQIYKNNKNEIEFKNIDNVQKEINKQDELAKLLDESKEFTKKQ 232
Qy 263 NFSKFTLGDMEML-----DVEGVADIDPNYK--FNOLLIHNNALSSVLMGSHNGIEPKVS 316
Db 233 ELNKYIDTKQEBELIKKLNDEKFNFNIDEKQKLLDQINSKINTLNENIKVMMLYETK--290
Qy 317 LLYGGNGPGARHDWNAVYGVKDOQGNVATIYVHMKNKSGLVGAGGEGKINN--PSFYL 375
Db 291 -----NKISNLQNEILNKDSTIKSLDEKQ---KLLDELKNNITSLYN 332
Qy 376 YKEDLTGTSQALSQEEIQNKIDFMFEFLAQNNAKLDNLSE--KEKEK--FTEIK-----DQ 429
Db 333 KSNITAIITQILLE-----SSLTDF-----NNANI--NINELSKTKLEFNDIOKLNNDIT 381
Qy 430 KDSKAYLDALGNDRIAFYKSKDKTHSALITEFGNGDLSYTLKDYKKADKALDRKENVTL 489
Db 382 EQNNKITDFFNSTRIPEKLDTEYKK--IDDIKNNNLOKLEESY--KKIDQETEYKNNKI--438
Qy 490 QGSLKHGDMFVDYSNFKYTNAKSNPNKGVGTNGVSHLEVGFNKV--AIFNLPDLNLLA 547
Db 439 -----NKEYNDIIEELKNNLQKLEENKIDQETEYKNNKINKEYNDIIEELKN--NNLQ 490
Qy 548 ITSFVVRNLEDKLTGKLSPOEANKLIDFLSSNKLVEKTLNF-----NKAVADAKNTGN 603
Db 491 KLEENKNNIDKLTK-----LNKDIESNTLFPNK--LNISDFKDKSREIAKLNTE 538
Qy 604 YDEVKKAQKLEKSLRKEHLEKEVEKKLES-----KSGNNKNWE-----AKAANSOK 652
Db 539 YEQLR---KDLLENKNTNLMKLDNKLSSLEQLYDSKKNILDDGIDKTYNSLSEKNDKI 595
Qy 653 DEIF-----ALINKEANRARAIAIYAQNKLGIKRELSKLENNVN 691
Db 596 DEYFSNIEKFDIYNVNIENKFIGNLDLSIINKIINDDQ-----PREYINSKIDS 644
Qy 692 KNLKDFDKSDFEFKNGKNDKFSKAEETLKALKGSVKOLG---INPEWISK-----738
Db 645 NELSTM---FDDIFNAKQO--IASITNNIENSNKIKDNEFIISNEDSSKLELDEIRKYK 700
Qy 739 -----VENLNAALNEFKNGKNDKFSKVTQAKSDLENSVKQVIIINQKVTDKVDNLNOAVS 792
Db 701 QQFDKIKDAMTEVSKFENTLOKDIDSISKSNINELTNAYD--LIINTKANDLDDKLNNGYS 758
Qy 793 VAKATGDFSVEQALADLKNF--SKEQLAQQAQKNESLNARKKSEIYQSVKNGVNTLVG 850
Db 759 -----EFKNLYNNASDLDDTIQKNDEKVKQOLNEYLEKNNK-----794
Qy 851 NCLSQAEATTLKSNFSDIKELNAKLGNFNNNNNGLNNEPIYAKVNNKKKQAQASLEEP 910
Db 795 -----QSEINDIVNNFIKELIFNNFTETNKSNEELL-----826

Qy 911 IYAAVKVNAKIDRLNQLASGLGVGAAGFPLKRRHKVDLDSKVLGRNQLAOKIDN 970
Db 827 ---TNDINDIKIFKL-----YKELNKISTNNLLKTYKNE-----IDN 860
Qy 971 LQAVSEAKAGFFGNLEOTIDKLKSTKHNPMNLWVESAKKVPASLSAKLDNYATNSHIR 1030
Db 861 VNEKLS-----IVENLQFINSLSEFFNOGTSITSHN 893
Qy 1031 INSNIKNG-----AINEKATGMITQKNPEWLK-LVNDKIIVAHNVGVSPLSEYDKIGFNQ 1083
Db 894 FLLNTLAGINDVNLKLNKIMADTTRGDTNIRDEIKNOISSENKISQK-----PNE 945
Qy 1084 KMKDYSDFKFSFKLNNVAVKDTSGTQFLTNAFSTASYCLARENAE-----HGI 1135
Db 946 KNEKDLKLISFNKLNK--YNISAGYTEYNNIEHCKLYLYIAVSDQERYRYRFTIHN 1003
Qy 1136 KNVN 1139
Db 1004 KQID 1007

RESULT 8
G70163
hypothetical protein BB0512 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: G70163
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: G70163
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2166 <KLE>
A:Cross-references: GB:AE001153; GB:AB000783; NID:g2688419; PIDN:AAC66876.1; PID:g268842
A:Experimental source: strain B31

Query Match 5.5%; Score 322.5; DB 2; Length 2166;
Best Local Similarity 19.1%; Pred. No. 0.00024;
Matches 259; Conservative 246; Mismatches 496; Indels 357; Gaps 63;

Qy 6 IDQOPQTEAENPQ--QFINNLQVAFKVDNA-----VAS---YDPQKPIVDKND 51
Db 129 VDKLSKTLKGFMTQIDSVESNLNVLEKFKDKANKENLESIKASWEKFDTNIKELVFKID 188
Qy 52 RDNROAF---EGISOLREYSNAIKNPTKKNQVFSDFPINKSNDLNKLDIDVESPFS 108
Db 189 NLNEISLYKDLANIERKKNDILVKGNEKLDLDFESDFLEKVENICKYSK-EIESSF-- 245
Qy 109 FQKFGDORYIFTSWVS--HQNDPSKINR-----SIRFMENIIQPIILDDEK 156
Db 246 --NFVENYKLIENSIELIMESVKNKINEKEDFTLNRLNEELQNKFKDIL--VYVDDRSK 301
Qy 157 A-----EFLKSAQSFAGIIG--NQIRTDQKF-MGVFDES-----KERQE 195
Db 302 EIKDKLEKLVLDVNETSSSMSSSKDNVYSRINSLESRMETMGKYESQVDVDFDKFRSQ 361
Qy 196 AEKNEPTGGDWLDIFLSFIEDKKQSDVKEAINOEPVPHVQPDITATTTDIOGLPPPEAR 255
Db 362 VELNLKNIYEDYED-----KISOVDNNIRE-----RVLSLLDLSNMESVQSACI 407
Qy 256 DL---LDERGN-----FSKFTLGDMEMLDVE---GVADIDPN-----YK 288
Db 408 DFKRLDDSNIGIVLEFKGKFGADIEVFSEFSFGKDGINDLQKQWLESQDLVDSDNIQEKLIK 467
Qy 289 FNQLLIHN-----NALSIVLMGS-----HNGIEP----- 312

Db 468 LNDNLISNFEENGNNYNSLNDNINAKYATLPESLDSSSKFENQMESKYKSFTDKL 527
Qy 313 ---EKVSLLYGGNGGPGCARHDWNAIVGYKDOQG-----NNVATIIINVHKN----- 355
Db 528 TAGMDEFSLMYGEKFETLSQ---EATNNYQFQDLNKLKLENEIESFYNNFEKTQETLKVD 584
Qy 356 -GSGLVITAGGEGKINNPSFYLYKEDOLTSQORALSQOEIQNKIDPFMEFLAQNNAKLDNLS 414
Db 585 FNTSLINIKDEIGNIVEF---RDRIYDEYVNIFTQLEESKLQYSKWQGEWDSNLKAKIE 640
Qy 415 ---EKEKEKTEIKDFQDKSKAYL-DALGNDRIAFVSKK-----DTFKHSAL 457
Db 641 SQINKTNEEFLSLIQ-IQDKGIELSESVFNDSDHIQKALDMHGSWKDELIALNKSLL 699
Qy 458 ITEFGNDL--SYTLKDYKKKADKALDREKVVTLQ-GSLKHDGVNMFVDSYFNKYTNASKN 514
Db 700 DIKVSSELLSSATLKIESLEK-VNDRMEYVLLKTGDI--ESLVIERYKELKDMYSQS 756
Qy 515 PNKGVTNGVSHLEVGFNKVAIFNLPLDNLNLAITSFVRRNLEDKLTTKGLSPQEAANKLI 574
Db 757 DEAILGIKEFINRQTEITIKDKSVFMLEDLN-----KXFDKKNFVSKIEEDCYKL 807
Qy 575 KDFLSSNKELVGKTLNFN-----KAVADAKNTGNY---DEVKKAQDKLESLRK 620
Db 808 KDFKIESEDILN---NFKSDLNEFIESKLQIYVSNIKSNQKQIDFLDRISKDI---LNR 861
Qy 621 REHLEKEVEKKESKGNKNKMEAKQ--ANSQKDEIFALINKEANRARAIAAYA-QNLK 677
Db 862 KDSINNEVDSLKSLDWOSKNEITVKIENLLSSGKVDL-DLIDSEVTTKIKELKFSIESLE 920
Qy 678 GIKRELSKLENNVKNLKDPKSDEFKNGKNGKDFSKAEETLKALKGSVKDLGINPEWIS 737
Db 921 SYYLEKIDEFNRQAGIYS--DELLQDINMHFNKETRELEENLSKKFAAV--LNNSEEVK 976
Qy 738 KV-----ENLNAALNEFGKNGKDFSKV-----TQAKSD 767
Db 977 EVDLSLQDKRTDIASFOANIDITLDSLVNKFNDINKEINGKYNEVISNVRGSENISSKL 1036
Qy 768 ENSVKDVI--INQVTKVDNQLNQA---VSVAKATGDFSRVEQALADLNKFSKQLAAQ 821
Db 1037 ENEMHETENLSRRLTDRIDSLSGMDENLOKLKESPDVSKYQVEKFEK-----VKDL 1090
Qy 822 AOKNESLWARKKSEIYQSVKNGVNGTL-----VGNGLSQAEATTLSKNFSDIKKELNAKL 876
Db 1091 TDDGEAKTNKLVKETEYQYKSRLEAIDYRRTIDNDIMQAK-----ERFGEITNELKNNI 1145
Qy 877 GNFNNNNGLN-----EPIYAKVKKKAGQAASLEEPIYAQVA---KKNNAKID 924
Db 1146 ESKSEFLNDLYKERFKLIESNFEEYSTFLIESEGAISKIRDEIYKTLTSDNENLQIKIS 1205
Qy 925 RLNQIASGLGVVGAAGFPLKRRHKVDLDSKVLGRNQLAOKIDN---LNOAVSEAKA 980
Db 1206 EMDQ-----NFEILEQSKDILEFEKELQDKICYCFINSQFGEIKA 1248
Qy 981 GFGNLEOTIDKLKDSKTHNPMNLWVESAKKVPASLSAKLDNYATNSHIRNS----- 1033
Db 1249 GVEENIKNHFD-----VCIKKVTNLDIDIVKYENEIHKRIDSLSKSIEST 1293
Qy 1034 --NLKNGAINKATCML---TQKNPEWLK-----VNDKIVAHNVGVSPL 1073
Db 1294 FDSIEKN-LNDKVSICIDKIANDFNLKYLEEERCNEGOLNLENKIDNKIKR--IDNLAL 1350
Qy 1074 SEYDKIGFNQNMKDYSDSF-KFSTKLNNAVKDTNSGF 1110
Db 1351 SOYDGL-----EKYADMYDEFSERLNSYIATLSEEF 1382

RESULT 9
T30822
lmp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 07-Dec-1999
C:Accession: T30822

R.Jensen, L.T.: Ladefoged, S.; Birkelund, S.; Christiansen, G.

Infect. Immun. 63, 3336-3347, 1995

A:Title: Selection of Mycoplasma hominis PG21 deletion mutants by cultivation in the pre

A:Reference number: 218884; MUID:95369882

A:Accession: T30822

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1365 <JEN>

A:Cross-references: EMBL:U21962; NID:G790243; PID:G790244; PIDN:AAA81013.1

C:Genetics:

A:Gene: Imp1

A:Genetic code: SGC3

Query Match 5.5%; Score 321; DB 2; Length 1365;

Best Local Similarity 22.3%; Pred. No. 0.00016;

Matches 244; Conservative 165; Mismatches 413; Indels 274; Gaps 58;

QY 58 FEGISOLREYENKAIKNPTKKNQYFSDFINKSNLDINKNDLIDVESSTKSFQKFGDORY 117

DB 124 FNEIKDKLOEY----IKNELSKOEYEHKQNTENEL-NKYTPISLESTLIEIQATNNLI 178

QY 118 RIFTSWSHONDPSKINTSRNEMENIOP----PILDDKEKAEFLKSAKQSFAGIIG 173

DB 179 KLLNESTREKONIDNLNAK--EQLKASISQANQLLPQSLDND-SEIAKAKSLDAEIKNA 235

QY 174 NOIRTDQPMGVDFDESLEKROPAEKNGEPTGDMWIDFLSFYDFKK-QSSDVKEAINEQEP 232

DB 236 NO-----AVASNTASMQSAKSLDAKVAEITKLETENKDKAEKAFNELKQTRNQ-- 285

QY 233 VPHVQPDIAITTTDIOGLPPPEARLDLDE-----RGNFSEFTIGDMELVGVADID-PNY 287

DB 286 ---IQEFINTNKN-----PNYSELISQTSKRDSKNSVT-----DSSNKSDESANT 330

QY 288 KPNQLLIHNLSSVLGSHNGIEPEKVSLLYGGNGGPGCARHDMNATVGYKDOQGNVAT 347

DB 331 ELKQALAKANA-----DKVQA-----DMLAKSIKEQLNNSVSN 363

QY 348 IINVMHNGSLVIAAGGEGGINPFIYLYKE-----DQLTGSQLALSOBEIQNKIDFMBFL 403

DB 364 ANTLISAK-----LTDKNTIQAKTELEKEVQKADQAIKSNNTASMQSAKSLDAK--V 415

QY 404 AQNNAKLONLSKEKEF-----RTEIKDF---OKDSKAYLDALGNDRIFAVSKKDTKH 454

DB 416 ABITTKLETENKDKAEKAFNELKQTRNQIOEFINTNKNPNYSELISQ-----LTSKRDSKN 471

QY 455 SALITEFGN-GDLSYTLKDYGKADKA-LDREKNVTLQGLSKHDGVMFVDYSNFKYTNAS 512

DB 472 S--VTDSSNKSDESANTELKQALAKANADKQVQADNLAKSIKEQLNNSVSNAN---TUSA 526

QY 513 KPNKGVGVTVNGVSHLEGVGNKVA--AIFNLPLNLAITSFVRNLEDKLT--TKGLSPQ 568

DB 527 KLTDKNTIQAKTELEKEVQKADQAIKS-----NNTASMQSAKSLDAKVAEITKKLETF 582

QY 569 EANKLIK-DFLSSNKELVGTILNPNKAVADAKNTGNYDEVKKAQKDKLSRKREHLEKE 627

DB 583 NKDKAEKAFNELKQTRNQIOEFINTN-----KNPNYSEL-----ISQTSKRD----- 625

QY 628 VEKLESKGNKKEAKAQAQNSOKDEIFALINKENARDAIAAONL-KGIKRELSDK 686

DB 626 -SKNSVTDSSNKSDESANTELKQALA-----KANADK---VOADNLAKSIKEQLNNS 674

QY 687 LENVN---KNLKDFDKSFDEKNGKNDKDFSKAEETLK-----ALKGSVKOLGIN-PEWIS 737

DB 675 VSNANTLSAKLTDKNTIQAKTELEKEIQKANOAIKSNNTASMQSAKSLDAKVAEITK 734

QY 738 KVENLN-----AALNEFK-----NGKNKDFSKVTO---AKSDLENSVKDVIINQ 778

DB 735 KLETENKDKAEKAFNELKQTRNQIOEFINTNKNPNYSELISQTSKRDSKNSVTDSNKS 794

QY 779 KVTDKVDNLNQAQVSAKATDFSRVEQALADLNKFSQEQALQAQAKNESLNARKKSELYQ 838

DB 795 DIESANTELKQALATAKAK--SSIDNELRPLKN-----DLQSKIIEFG 836

QY 839 SYKNGVNGTFLVGNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNPEIYAKVNK 898

DB 837 PIRN-TNFSWISSKLE---TTKNLAELTKA-----DAKNNPSSSKQAL 878

QY 899 KKAGQAASLEPEIYAQVAKKVNAKIDRLNQIASGLGVQQAAGFPLKRHKDVDDLKSKVGL 958

DB 879 KDSSQ-----QVOKLGN---ELLKTIIEFG-----KVE----- 904

QY 959 SRNQELAKIDNLNQAQVSEAKAGFEGNLEQTTDKLKDSTKHNPMNLNWSAKKVPASLSA 1018

DB 905 TKNSNIGYRFLKLAQAEQ-----FNNSD--VDKLLK-----NAWEE--KOTLLSKKQ 946

QY 1019 KLDNYATNSHIRINSNIKNGAINERKATMLTQKNPEWLKLVNDKIVAH----NVGSPVLS 1075

DB 947 KLGNSTKDY-----LTOLSTEMSTQEST--IKKVIVNIQAHIRNLSQYRL 993

QY 1076 YDKIGFNOKMKDYSD 1091

DB 994 ADKLIANMK--RGYGD 1007

RESULT 10

S67593

transport protein USO1 - Yeast (Saccharomyces cerevisiae)

N:Alternate names: protein D2552; protein YDL058w

C:Species: Saccharomyces cerevisiae

C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000

C:Accession: S67593; A38455; S30782

R:Blöcker, H.; Brandt, P.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67587

A:Accession: S67593

A:Molecule type: DNA

A:Residues: 1-1790 <BLO>

A:Cross-references: EMBL:Z74106; NID:G1431058; PID:e253003; PID:G1431059; MIPS:YDL058

A:Experimental source: strain S288C

R:Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.

J. Cell Biol. 113, 245-260, 1991

A:Title: A cytoskeleton-related gene, USO1, is required for intracellular protein tra

A:Reference number: A38455; MUID:91185402

A:Accession: A38455

A:Molecule type: DNA

A:Residues: 1-389, 'TA', 392-724, 'S', 726-1790 <NAK>

A:Cross-references: GB:X54378; NID:G4777; PIDN:CAA38253.1; PID:G4778

A:Note: the authors translated the codon ACT for residue 768 as Ile

R:Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.

submitted to the EMBL Data Library, February 1993

A:Description: An integrin analogue in Saccharomyces cerevisiae.

A:Reference number: S30782

A:Accession: S30782

A:Molecule type: DNA

A:Residues: 71-846, 'E', 848-923, 'K', 925-1252, 'I', 1254-1318, 'V', 1320-1460, 'S', 1462-1580

A:Cross-references: EMBL:L03188

C:Genetics:

A:Gene: SGD:USO1; INT1

A:Cross-references: SGD:S0002216; MIPS:YDL058w

A:Map position: 4L

C:Keywords: coiled coil; transmembrane protein

F:326-342/Domain: transmembrane #status predicted <TM1>

F:394-410/Domain: transmembrane #status predicted <TM2>

F:617-633/Domain: transmembrane #status predicted <TM3>

Query Match 5.4%; Score 317; DB 2; Length 1790;

Best Local Similarity 18.5%; Pred. No. 0.0003;

Matches 233; Conservative 213; Mismatches 467; Indels 344; Gaps 48;

QY 3 NETIDQQPOTAEAFNPQOFINNQLQVAFLLKVDN-----AVASVDPDOKPIVDKNDNR 55

DB 677 NEDSILPELDETGLPKVYFSTY---FIQLFNENIYRTRTALSHDPDEEPI-----NK 726

QY 56 QAFEGISQLR-----EYSNKN--AIKNPTKKNQYFSDFINKSND 92

Db 508 ITNIEGALKESKGNYEIGF-----LEKLEETG-----KNRKLKVDITKKSI-----NST 551
QY 574 IKDFLSSNRELKVGKTLNFKAVADAKNTGNYDEVKKAQKDLEKSLR-REHLEKEVEKKL 632
Db 552 VGNFSSLENNFDLNOYDFKNINDYEN-----KMGIIYNEFEGLSNKISENLNASENTS 606
QY 633 ESKSGNKNMEAKQAANSOKDEIFALINK--EANDARAIAYAQNLKIGIKRELSDKLENV 690
Db 607 DYNASAKTLREA-----OKEKV-NLLNKEEANKYLROVKKVESPREI-FNNKESLDKI 658
QY 691 NNKLDFDKSFDEFKNGKNKDKFSKABETLALKGSKVGLGGINPEWISKVENLNAALNEFK 750
Db 659 NEMIK-----KEQITVNEGHGVNKVQJENIKELVDENNLSILQO-A 699
QY 751 NGKNKDFSKVTOAKSDLENSVKDVI-----INQKVTD----- 782
Db 700 TCKNEISQKITH--STLKNKAKTILGHVDTSAKYVGIKIPPELALTTELCDAKLTAQEL 757
QY 783 KVDNLNOAV-----SVAKATGDFSRVQALADLNKFSKEOLAQO-----AQKNESLNARKK 833
Db 758 KPESKNNVLETEMMSKNTNELD-VHKNIQDAYKVALEILAHSEIDTDTKODKSSKLIENG 816
QY 834 SEIYQSV-----KNCVNGTLVNGLSQAETLTKSNFSDIKK-----ELN 873
Db 817 NQIYLKVLINQYKKNKISSIKSKEEAUVKIGNVSKKHSKLSKITCSDKSYDNIILAEKQ 876
QY 874 AKLGNENN-----NNNN-----GLKN--EPIYAKVNNKKAG-----QAASL 907
Db 877 TELONLRNSFTQKNTWNSDSKLEIKTDFESLKNALKTLEGEVNAKASSONHEHVQSK 936
QY 908 EPIYAQVA--KKVNAKIDRLNQIASGLGVGQAGFPLKRHRDKVDLSDSKVGLSRNOELA 965
Db 937 SEPVNPALSEIEKEETDIDSLNTALDELKKGRTG--EVSRYKLKIDVTYKELSDDEL- 993
QY 966 QKIDNLNOAVSAKAGFFNLEQTDKDL-----KOSTHNPMLNWSAKKVPASL 1016
Db 994 --INTIEKVVKAYLAIYIKKNYEDTVODVLTLEHFNFTKQVSNHEPTNF--DRSKN-----S 1045
QY 1017 SAKLDNYATNSH-----IRNSNIKNGAINEKATGMLTQKNPEWLKLVNDKIVAH 1066
Db 1046 SEELKAVYDTSKTIISKLGKVIIEVNEENTMTIESSAKEIEALYN-----ELKNKTSUN 1101
QY 1067 NV-----GSVPLSEYDKIGFNQKNKDYSDSPKFTSKLNNVADTNSGFTQFLTNAFSTAS 1122
Db 1102 EIYQTSNEVKLOE-----MKSADKYIDVSKIFNTVLDTK--SNIVTN----- 1143
QY 1123 YCLARENAEHGKNVNTK 1141
Db 1144 -----OHSINNVDK 1153

RESULT 12

myosin heavy chain - Entamoeba histolytica
C:Species: Entamoeba histolytica

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000

C:Accession: T18296

R:Guillen, N

submitted to the EMBL Data Library, February 1997

A:Reference number: 218865

A:Accession: T18296

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2139 <GUI>

A:Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1

C:Genetics:

A:Gene: mhca

C:Superfamily: myosin heavy chain; myosin motor domain homology

F:91-780/Domain: myosin motor domain homology <MOW>

Query Match

Best Local Similarity 5.1%; Score 302; DB 2; Length 2139;

Pred. No. 0.0013;

Matches 249; Conservative 206; Mismatches 429; Indels 354; Gaps 54;
QY 59 EGISQRLREYSNKAIKNPTKKQYF--SDFINK--SNDLINKNDLID--VESSTKFSQ--- 110
Db 706 EGIRTRMCPYPRMIYSEFLKRYFLGDKINRRASDEKATASTLMRQIIEIKYKDKQCPRA 765
QY 111 KFGDORYIFTSWVSHQNDPSKINTR-SIRNFMENIQPPIILDDKEKABFLKSAKOSFAG 169
Db 766 KEGDYRF-----CKTKIFFRVGILAEIEEVREAKI--SEMLGGLQACARAFAA 811
QY 170 IIGNQIRTDQKPMGVFDESLEK-----ERQEAENKGEPTGGDWL 208
Db 812 RKHFSVLRBQQAVIVIQRTLKSWTFRNWMWMLYVKARPMLKRNFEEGKKDKKDETE 871
QY 209 DIFLSFIDFKOSSDVKEAINOEPPVHPVQPDIAITTTDIOGLPPEARDLDDERGNF-SKF 267
Db 872 DL-----KKLAEEIKKREAAENA-----LASATAKTGELEAKIQDLEDEKISELESKL 919
QY 268 TLGDMEM-----LDVEGVADIDPNYFNQOLLIHNNALSSVLMGSHNGIEPEKVSLLYGGNG 323
Db 920 SAAELDKQELNLKIEENLEEDKEELKETIDNLKGLDKSLKGEDLEVEITELN----- 972
QY 324 GPGARHDNNATVGYKDDQCGNNVATTINVHMKNKSGLVIAAGGKGINNPSFYLYKEDOLT- 382
Db 973 --SQINTLNATYDKKDK-----TIAEMQESIDE-----KEDEITK 1005
QY 383 --GSORALSOEIQNKIDPFMEFLAQNNAKLDNLSEKEKEKFEITEIKDFQKSKAYLDALG 440
Db 1006 LKGDIKLLEEE-----KDDLEQDRADVSA TKDDIA-KLKNKTIIECED-----A 1048
QY 441 NDRIAFVSKKDKTSHALITEFGNGDLSYTLKDYGKKADKALDREKNVTLQGLSLKHGVMF 500
Db 1049 KDEIA-----KLEGELEDEEN----- 1064
QY 501 VDYSEKTYTNASKNPNKGVGTNGVSHLEVGNKVAIFNLPDLNLAITSFVRRNLE-DK 559
Db 1065 ---KNKDLTNELQOTQLGETEKSAAQVAATKASDERDTLS-----QNLNENK 1112
QY 560 LTTKGLSQEANKLIKDFLSSNKLGVGTNLNFKAVADAKNTGNYDEVKKAQKDLEKSLR 619
Db 1113 LTTKNTLTAKD-----LEKKISGLKQDY-EDLEDDKN-----KIEGDLRNAQR 1155
QY 620 KREHLEKEVEKKLE-SKSGNKNMEAKQAANSOKDEIFALINKAENRDRARAYAQNLKAG 678
Db 1156 KIKELDDEITKADYVSOYLQOKOKEEVESQIAKQBEKEAIGNDVKNKEKTIKEKELEIQS 1215
QY 679 IKRELSO---KLENVNKLKDFDK-----SFDEPKNGKNKDFSKAETLALKAGS 725
Db 1216 LOEKLDETEVEKEDAEKKKKEIEKEMKALQEBKENVESSEKSTEDKKKLEDNLKDTQKK 1275
QY 726 VKDLGINPEWI-SKVENLNAALNEFKNGKNK-----DFSKV 760
Db 1276 LDDMTADNEKLRKAKADLEAQLNEVODNHEKAVADAELLNKKKAQSDKELNSLKAELEAL 1335
QY 761 TOAKSDLENSVKDVIINOK--VTDKVDNLNOAV-----SVAKATGDFSRV----- 803
Db 1336 TRAKSVVESKND-SENEKAALSEIDQANEKLNQIADLRKATADLOAEANEKAEVEAQ 1394
QY 804 -EQALADLNKFSK--EQL-----AQQAOKNESLNARKKSEI--- 836
Db 1395 RDKLVADNKKMTKTLLEEIKARDEENTYKVENYKVLKRKEADLEENANLDEKDRMKNK 1454
QY 837 -----YQSVKNGVNGTL-----VGNGLSQAEATTLK-----NFSDIKKELNAKL 876
Db 1455 EKQVKKLGELEKTKDKKLNAAIAEKDSIPTAKKKQSDADLEELNKVTVEEHDEVVAKLNTQI 1514
QY 877 GNFNNNNNGLK--NEPIYAKVNNKKAGQAASLEEPIYAQVAKV-----NA 921
Db 1515 TKLTRDNQSAEELNE-LRSKADKDK-KISELEQVNELESRRPVGTGNADENEIKIRDA 1572
QY 922 KIDRLNQIASGLGVG---QAAGFFPLKRHRDKVDDLSKVGLSRNQ-----ELAOKTD 969
Db 1573 QIADLNKALEMKGVQNNQLOATNKEKAKDN-DLTSKIEITENEMKKLENKAKRLEQDKD 1631

submitted to the EMBL Data Library, November 1998

A:Reference number: Z18937

A:Accession: T18477

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2523 <LAW>

A:Cross-references: EMBL:AL008970; NID:e1407852; PID:e1322571; PIDN:CAAL5620.1

C:Genetics:

A:Map position: 3

A:Introns: 148/3

A:Note: C0485W

Query Match 5.0%; Score 294.5; DB 2; Length 2523;
Best Local Similarity 21.6%; Pred. No. 0.001;
Matches 258; Conservative 177; Mismatches 443; Indels 341; Gaps 61;

QY 30 LKVDNAVASYDP----DQ-KPIVDKNDNRQAFEGISOLREYSNKAINKPTKKNQVFS 84
DB 982 LKGENEI-GMDPLMKIDQTNKIVSKVDSNFKVNGDGINFNKTDGSGN-----FNKIDG 1032
QY 85 DFINKSNDL-INKNDLIDVESSTK-SFOKFGDQRYRIFTSWVSHQND--PSKINTSRIN 140
DB 1033 INFNTDGINFNKTDGINFNKTDGINFNKVDNINFNKIDKEVEKYVDPLPUSHRTDDMR- 1091
QY 141 FMENIQQPILDDEKAEFLSAKSAFAGIIGNQIRTDQKFMGVFDESLEKQERQAEKNG 200
DB 1092 -----KKKSELLSKDGS---IISN-----LDTSHFE----- 1116
QY 201 EPTGGDWLDIFLSFIFDK--KOSSDVKEAINOEPPHVQPDIAITTTDIOGLPPPEARDLL 258
DB 1117 -----INLSRSETONEMCKNSVKQLENKLEKEIKDEKELQ-----NEL- 1162
QY 259 DERGNFSKPTLGDMEMLDVEGVADIDPNVKFNOLLIHNNALSSVLMSHGSHGIEPEKVSLL 318
DB 1163 -ERSNWS-----IDIE---DLDKD-----LIINKESRDIKYKIWIIDINKDNYMMI 1203
QY 319 YGNGGPGARHDWNAVTVGYKQGGNVATII---NVHMKNGSLVIAGGEKINNPSFYL 375
DB 1204 YOON-----KCGRRKMKISQNKLLIKKRIKRNHE-----KKRKIRFFPEKL 1245
QY 376 YKEDQLTGSORALSQOEIQNKIDFMEFLAQNNAKLDNLSEKEKEFRTEIKDFQKDSK-- 433
DB 1246 YK-----RNDTHKKLRPIRFVRHVDVKLDNLNDK--TVMLKNEIRVDKGEDEGE 1292
QY 434 -AYLDALGNDRIA--FVSKKOTKHSALITEFGNGDLSYTLKYDGKKADKALDREKNVTQ 490
DB 1293 DYVEFDLKNKNNMGNMKNKKNVKNVKNVNN-----VKDV-----KNVNVKNVNV 1341
QY 491 GSKLDGGMVFDVSNFKYT---NASKNPNGVGVTVNGVSHLEVGNKVAIFNLPLD----- 543
DB 1342 NNVN-----VKDVKNMEHIDYKNKKEVMIKKKGESNNVPHEKHNKNKKNYCNV-DLGMHS 1396
QY 544 --NLAITS-----FVRNLEDKLTGSLSPQEAANKLIKDFLSSNKELVGKTLNFKAVA 596
DB 1397 LQNRHITITSEVSSKFLCKRMKNYF--DKSNNSIEIHKISASNIFRITMVCVASNIK-----G 1450
QY 597 DAKNTGNYDEV-----KKAQKDLKSLRKRHELEKEVEKLESKSGNKN 641
DB 1451 ENKNGNINNYKGPATKALVKNLFLFSKESKRAITSSKKRDDNNINVIKINTPS---QK 1507
QY 642 MEAKAANSQKDEIPALLNKEANRARAIAQNLIK--GIKRELSDKLENVKNLKD--- 696
DB 1508 VSEKRNNNNNNNVLGDKNKNKNDE---LFTKEIKKSTISQKQKKGNEGNTKTHKDNIN 1564
QY 697 -FDKSFDEF-----KNGKNKDFSKAETLALKGSKVDLGINPEWISKVENLNA 744
DB 1565 ILNEDVDHFQPSLRLEVTYKNNKN-----KNNKNKKFNDNTNNNNNNNS 1612
QY 745 ALNEFKNGKNKDFSK---VTOAKSDL-----ENSVKD--VIINQKVTDKVDNLNQAIVSAK 795
DB 1613 --NDPEEYKEEHIAITNEIVRESESDLYVSCDEGCVKNGDIYNMEIINNVDNIHK---IDK 1667

QY 796 ATGDFSRVQALADLKNSKEQL-----AQAQAKNESLNARKKSEIYQSVKNG 843
DB 1668 KGNDIDYKDKS---LENNKKQKGLIKLTPSSYTEKEKEKEKERRKKNNIYIP--- 1720
QY 844 VNGTLVGNCLSOAEATITLSKNFSDIKKELNAKLGNFNNNNNGLNNEPIYAKV---NKK 899
DB 1721 -----LTIASRTVAIPIINDTTKDVKNKLHLTKRNTC-----VTYCNVDNIQNK 1765
QY 900 KAGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVVGOAGFPKLRHDKVDLDSKVGLS 959
DB 1766 KKGD-----DKKNIKRQHVGLKLEFDEM--SAMFEKKKKIKKDDINKKEDI 1810
QY 960 RQELAQIDNLNQAVSEAKAGFFGNLEQTIDKLDKSTKHNPMLNWLWESAKKVPASLSAK 1019
DB 1811 NKDDINKKDNINK-----KNDINKKDDINKKDDINKKDDINKKDDIN---K 1854
QY 1020 LDNYATNSHIRINSNI-----KNGAINEKATGMLT-----QKNPEWLKLVNDKIYVAH 1066
DB 1855 KNNYNNNSN---NNNVVKFSKTHQNEEKIKGNITVIRNKLKDKGKENIGLKKKIERK 1911
QY 1067 NVGSVPLSEYDK-IGFNOKNMKDYSDSFKFSKTLNNAVKDTNSGFTQFLTNFASSTASYC 1125
DB 1912 NTTTTATKHDNIIDIKKKEKE-----NKMKN-----SKFQC 1945
QY 1126 LARENAE-HGINKVNTKGG 1143
DB 1946 LNKGTQIENKKNMIISG 1964
RESULT 15
A:64465
hypothetical protein MJ1322 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: A64465
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Black, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, A.; Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1005 <BUL>
A:Cross-references: GB:U67572; GB:L77117; NID:g1591958; PIDN:AAB99331.1; PID:g1591962
C:Genetics:
A:Map position: REV1273394-1270377
C:Superfamily: hypothetical protein MJ1322

Query Match 5.0%; Score 293.5; DB 2; Length 1005;
Best Local Similarity 21.6%; Pred. No. 0.001;
Matches 183; Conservative 134; Mismatches 278; Indels 251; Gaps 41;

QY 338 KDOQGNVATTIIVHMKNGSLVIAGGEKGINNPSFYLYKEDQLTGSQRALSQBEIQNKI 397
DB 6 KEIRNNPKSHVNSRIKEKGIIVGSGKSSIF-----EAV 45
QY 398 DFMEFLAQNNAKLDNLSEKEKEFRTEIKDFQKDSKAYLDALGNDRIFAIVSKKDPKHSAL 457
DB 46 PFALFGAGSNFYDTIITKGGKVVVEL--DPEVAGNNY-----KI 84
QY 458 ITERGNDLSYTLKYDGKKADKALDREKNVTLOGSLKHGDMVFDVSNFK-YTNASKNPN 516
DB 85 IREYDSG-----RGAKL-----YKNGKPYATTISAVN 112
QY 517 KGVGVTVNGVSHLEVGNKVAIFNLPLDNLNLAITSFVRNLEDKLTTKGLSPQEAANKLIK 576
DB 113 KAVNEILGVD-----RNMFLNSIYIKQGE---IAKFLSLKPSKEL--- 149
QY 577 FLSSNKELVGKTLNFKNAVADAKNTGNYDEVKKAQKDLKSLRKRHELEKEVEKLESKS 636

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2002, 14:02:21 ; Search time 68.44 Seconds

(without alignments)
614.473 Million cell updates/sec

Title: US-09-360-685A-5

Perfect score: 5867

Sequence: 1 MTNETIDQPPQTEAENFPOQ.....RENAERGIKNVNTKGGFQKS 1147

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5867	100.0	1147	1 CGAL_HELPY	P80200 helicobacte
2	5395.5	92.0	1186	1 CAGA_HELPY	P55980 helicobacte
3	5270.5	89.8	1182	1 CGA2_HELPY	P55746 helicobacte
4	5086	86.7	1167	1 CAGA_HELPJ	Q921t1 helicobacte
5	314	5.4	1790	1 US01_YEAST	P25386 saccharomyc
6	307	5.2	1251	1 RBP2_PLAVB	Q00799 plasmodium
7	301	5.1	1957	1 YB86_SCHPO	Q10411 schizosacch
8	291	5.0	2230	1 G0G4_HUMAN	Q13439 homo.sapien
9	289	4.9	2663	1 CENE_HUMAN	Q02224 homo.sapien
10	284	4.8	1875	1 MLPI_YEAST	Q02455 saccharomyc
11	279	4.8	1679	1 YI09_YEAST	P40457 saccharomyc
12	276.5	4.7	2869	1 RBP1_PLAVB	Q00798 plasmodium
13	276	4.7	1928	1 MYS1_YEAST	P08964 saccharomyc
14	272.5	4.6	2022	1 ANTI_ONCVO	P21249 onchocerca
15	268	4.6	1208	1 PCP1_SCHPO	Q92351 schizosacch
16	267	4.6	1727	1 ALMI_SCHPO	P47460 mycoplasma
17	267	4.6	1805	1 HMW2_MYCEE	Q9utk5 schizosacch
18	266	4.5	1938	1 MYS2_CAEEL	P02567 caenorhabdi
19	265	4.5	1966	1 MYSB_CAEEL	P02566 caenorhabdi
20	262	4.5	2116	1 MYS2_DICDI	P08799 dictyosteli
21	260.5	4.4	2704	1 BRA1_HUMAN	Q03001 homo.sapien
22	255	4.3	1935	1 MYH7_RAT	P02564 rattus.norv
23	254.5	4.3	1312	1 RA50_YEAST	P12753 saccharomyc
24	252	4.3	1938	1 MYH4_RABIT	P28641 oryctolagus
25	251.5	4.3	1959	1 MYH9_CHICK	P14105 gallus.gall
26	250	4.3	1972	1 MYHB_HUMAN	P35749 homo.sapien
27	249.5	4.3	1938	1 MYS_AEQIR	P24733 aequipecten
28	249	4.2	1630	1 MSP1_PLAFK	P04932 plasmodium
29	249	4.2	1639	1 MSP1_PLAFW	P04933 plasmodium
30	248.5	4.2	1972	1 MYHB_RABIT	P35748 oryctolagus
31	247.5	4.2	1972	1 MYHB_MOUSE	O08638 mus.musculu
32	247	4.2	1934	1 MYH7_MESAU	P13540 mesocricetu
33	246.5	4.2	1939	1 MYH1_HUMAN	P12882 homo.sapien

34	245.5	4.2	1939	1 MYH6_MESAU	P13539 mesocricetu
35	245	4.2	1935	1 MYH7_PIG	P79293 sus.scrofa
36	245	4.2	1938	1 MYH6_MOUSE	Q02566 mus.musculu
37	245	4.2	1939	1 MYH6_HUMAN	P13533 homo.sapien
38	243.5	4.2	2349	1 TPR_HUMAN	P12270 homo.sapien
39	243	4.1	1818	1 HMW2_MYCPN	P75471 mycoplasma
40	243	4.1	2104	1 MYS3_SCHPO	O14157 schizosacch
41	242	4.1	1427	1 REST_HUMAN	P30622 homo.sapien
42	240.5	4.1	1938	1 MYH6_RAT	P02563 rattus.norv
43	240	4.1	1935	1 MYH7_HUMAN	P12883 homo.sapien
44	238.5	4.1	1960	1 MYH9_HUMAN	P35579 homo.sapien
45	238.5	4.1	2710	1 TOXA_CLODI	P16154 clostridium

ALIGNMENTS

RESULT 1

ID	CGAL_HELPY	STANDARD;	PRT; 1147 AA.
AC	P80200;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN (120 KDA PROTEIN).		
GN	CAGA OR CAL.		
OS	Helicobacter pylori (Campylobacter pylori).		
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;		
OC	Helicobacter.		
OX	NCBI_TaxID=210;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CCUG 17874 / NCTC 11638;		
RX	MEDLINE=93296225; PubMed=8516329;		
RA	Covacci A., Censini S., Bugnoli M., Petracca R., Burrone D.,		
RA	Macchia G., Massone A., Papini E., Xiang Z., Figura N.,		
RA	Rappuoli R.;		
RT	"Molecular characterization of the 128-kDa immunodominant antigen of Helicobacter pylori associated with cytotoxicity and duodenal ulcer.";		
RT	Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN=CCUG 17874 / NCTC 11638;		
RX	MEDLINE=97121442; PubMed=8962108;		
RA	Censini S., Lange C., Xiang Z., Crabtree J., Ghiara P.,		
RA	Borodovsky M., Rappuoli R., Covacci A.;		
RT	"cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors.";		
RT	Proc. Natl. Acad. Sci. U.S.A. 93:14648-14653(1996).		
RL	[3]		
RP	SEQUENCE OF 181-90;314-28;366-77;420-30;494-506;661-77;900-14;1062-77.		
RA	Herrmann V., Herrmann J., Kist M.;		
RL	Submitted (APR-1993) to the SWISS-PROT data bank.		
CC	!- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT, OR FUNCTION OF THE CYTOTOXIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X70039; CAA49633.1; -		
DR	EMBL; AF282853; AAC44706.1; -		
DR	HSSP; P02937; IMLP.		
KW	DOMAIN	880	885
FT	CONFLICT	320	320
FT	CONFLICT	325	325
FT	CONFLICT	328	328
FT	CONFLICT	328	328
FT	POLY-ASN.		
FT	G -> A (IN REF. 3).		
FT	P -> F (IN REF. 3).		
FT	R -> K (IN REF. 3).		

```
FT CONFLICT 426 426 K -> E (IN REF. 3).
FT CONFLICT 429 429 AON -> TED (IN REF. 3).
FT CONFLICT 673 675 A -> T (IN REF. 3).
FT CONFLICT 901 901 A -> E (IN REF. 3).
FT CONFLICT 903 903 L -> P (IN REF. 3).
FT CONFLICT 907 907 L -> P (IN REF. 3).
FT CONFLICT 910 910 P -> R (IN REF. 3).
FT CONFLICT 914 914 P -> E (IN REF. 3).
FT CONFLICT 1072 1072 P -> S (IN REF. 3).
FT CONFLICT 1074 1074 S -> D (IN REF. 3).
SQ SEQUENCE 1147 AA; 128013 MW; AB92770835F68490 CRC64;

Query Match 100.0%; Score 5867; DB 1; Length 1147;
Best Local Similarity 100.0%; Pred. No. 5.5e-197;
Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTNETIDQOQTEAFNPQOFTNNLQVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOQTEAFNPQOFTNNLQVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
QY 61 ISOLREYSNKAIKNPTKKNQYFSPFINKSNDLKNKNDLIDVESSTKSPQKFCDDORYRIF 120
Db 61 ISOLREYSNKAIKNPTKKNQYFSPFINKSNDLKNKNDLIDVESSTKSPQKFCDDORYRIF 120
QY 121 TSWVSHQNDPSKINTSRNFMENIIQPPILDDKEAEFLKSAKOSFAGIIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTSRNFMENIIQPPILDDKEAEFLKSAKOSFAGIIIGNQIRTDQ 180
QY 181 KPMGVFDESLEKQBAEKNGEPTGGDWLDFLSFTFDKQSSDVKEAINQEPVPHVQPD 240
Db 181 KPMGVFDESLEKQBAEKNGEPTGGDWLDFLSFTFDKQSSDVKEAINQEPVPHVQPD 240
QY 241 ATTTTDIOQLPEARDLDERGNFSGFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
Db 241 ATTTTDIOQLPEARDLDERGNFSGFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGCARHDWNTATVGYKDOQGNVATIIINVHMKNGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGGPGCARHDWNTATVGYKDOQGNVATIIINVHMKNGSLV 360
QY 361 IAGGEGKINNPFFYLYKEDQLTGSORALSQSEIQNKIDPMEFLAONNAKLDNLSEKEPK 420
Db 361 IAGGEGKINNPFFYLYKEDQLTGSORALSQSEIQNKIDPMEFLAONNAKLDNLSEKEPK 420
QY 421 FRTKIDFQKSKAYLDALGNDRIFAIVSKKDKHSALITEFGNGDLSYTLKDYGGKADKA 480
Db 421 FRTKIDFQKSKAYLDALGNDRIFAIVSKKDKHSALITEFGNGDLSYTLKDYGGKADKA 480
QY 481 LDREKNVTIQQSLKHDGVNFVDYSNFKYTNASKNPNKGVGTNGVSHLEVGFKNVAIFNL 540
Db 481 LDREKNVTIQQSLKHDGVNFVDYSNFKYTNASKNPNKGVGTNGVSHLEVGFKNVAIFNL 540
QY 541 PDLNNLATSFRVNRNLEDKLTTKGLSPQBAKLIKDFLSSNKELVCKTLNPNKAVADAKN 600
Db 541 PDLNNLATSFRVNRNLEDKLTTKGLSPQBAKLIKDFLSSNKELVCKTLNPNKAVADAKN 600
QY 601 TGNVDEVKAQDLKSLKREHLEKEVEKLESGNKNKNKMEAKAQAQNSQKDEIFALIN 660
Db 601 TGNVDEVKAQDLKSLKREHLEKEVEKLESGNKNKNKMEAKAQAQNSQKDEIFALIN 660
QY 661 KEANDARAIAAQNKGKIKRELSKLENVKNLKDFOKSFDFKNGKNKDFSKAEETLK 720
Db 661 KEANDARAIAAQNKGKIKRELSKLENVKNLKDFOKSFDFKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENIENALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENIENALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
QY 781 TDVDNLNQAVSVAKATGDFSRVEQALADLNKFSKEQLAQQAQKNESLNARKKSEIYQSV 840
Db 781 TDVDNLNQAVSVAKATGDFSRVEQALADLNKFSKEQLAQQAQKNESLNARKKSEIYQSV 840
```

```
QY 841 KNGVNGTLVGNGLSQAEATTLTSKNFSDIKKELNAKLGNNNNNNGLKNEPIYAKVKKK 900
Db 841 KNGVNGTLVGNGLSQAEATTLTSKNFSDIKKELNAKLGNNNNNNGLKNEPIYAKVKKK 900
QY 901 AGQAASLEPIYAQVAKVNAKIDRLNQITASGLGVVQGAAGFPLKRHRDKVDLDSKVGLSR 960
Db 901 AGQAASLEPIYAQVAKVNAKIDRLNQITASGLGVVQGAAGFPLKRHRDKVDLDSKVGLSR 960
QY 961 NOELAQKIDNLNQAVSEAKAGFFGNLEQITDKLKDSTKHNPMLNVAVESAKKVPASLSAKL 1020
Db 961 NOELAQKIDNLNQAVSEAKAGFFGNLEQITDKLKDSTKHNPMLNVAVESAKKVPASLSAKL 1020
QY 1021 DNYATNSHIRINSNIKNGAINEKATGMLTKQKNPEWLKLVNDKIVAHNVGSVPLSEYDKIG 1080
Db 1021 DNYATNSHIRINSNIKNGAINEKATGMLTKQKNPEWLKLVNDKIVAHNVGSVPLSEYDKIG 1080
QY 1081 FNQKNMKDYSDSFKESTKLNNVAVKDTNSGFTQFLNAPSTASYCYCLARENABHGKKNVT 1140
Db 1081 FNQKNMKDYSDSFKESTKLNNVAVKDTNSGFTQFLNAPSTASYCYCLARENABHGKKNVT 1140
QY 1141 KGGFOKS 1147
Db 1141 KGGFOKS 1147

RESULT 2
CAGA_HELPY
ID CAGA_HELPY STANDARD; PRT; 1186 AA.
AC P55980;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN (120 KDA PROTEIN)
DE (CAG PATHOGENICITY ISLAND PROTEIN 26).
GN CAGA OR CAI OR CAG26 OR HP0547.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX Helicobacter.
OC NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori."
RL Nature 388:539-547(1997).
CC -!- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
OR FUNCTION OF THE CYTOTOXIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000569; AAD07614.1;
DR TIGR; HP0547;
KW Antigen; Complete proteome.
FT DOMAIN 247 250 POLY-THR.
FT DOMAIN 883 889 POLY-ASN.
SQ SEQUENCE 1186 AA; 132386 MW; B05C3F2CCC4444F4 CRC64;
```

Query Match 92.0%; Score 5395.5; DB 1; Length 1186;
Best Local Similarity 89.9%; Pred. No. 1.3e-180;
Matches 1068; Conservative 26; Mismatches 51; Indels 43; Gaps 5;

QY 1 MTNETIDQ-----QPTEAAFPQOFINNLOVAFKLVNDNAVASYDDPKPIVDKNDNR 55
DB 1 MTNETIDQTRTDDQVQSFQAFDPQOFINNLOVAFKLVNDNAVASFDDPKPIVDKNDNR 60

QY 56 QAFEGISQREYSNKAINKPTKKNQYFSDFINKSNDLINKNDLIDVESSTKSFQFGDQ 115
DB 61 QAFDGISQREYSNKAINKPTKKNQYFSDFINKSNDLINKNDLIDVESSTKSFQFGDQ 120

QY 116 RYRIFTSWVSHONDPSKINTRISIRNPMENIIPPLDDKEAEFLKSAKQSFAGIIGNQ 175
DB 121 RYQIFTSWVSHOKDPSKINTRISIRNPMENIIPPIPPDDKEAEFLKSAKQSFAGIIGNQ 180

QY 176 IRTDQKFMGVFDESILKEROEAKNGEPTGGDWLDFLSPFDKQSSDVKEAINOEPVPH 235
DB 181 IRTDQKFMGVFDESILKEROEAKNGEPTGGDWLDFLSPFNKQSSDVKEAINOEPVPH 240

QY 236 VQPDIAATTTTDDIQLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLTH 295
DB 241 VQPDIAATTTTDDIQLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLTH 300

QY 296 NNALSSVLMGSHNGIEPEKVSLLYGGNGPGARHDWNAVTVGYKDOQGNVATLINVHMKN 355
DB 301 NNALSSVLMGSHNGIEPEKVSLLYAGNGFGDGKHDWNAVTVGYKDOQGNVATLINVHMKN 360

QY 356 GSGLVITAGEKGINNPSFYLKREDOLTGSRALSOEIQNKIDFMEFLAQNNAKLNDLSE 415
DB 361 GSGLVITAGEKGINNPSFYLKREDOLTGSRALSOEIRKVDMEFLAQNNTKIDNLSE 420

QY 416 KEKEKFEITEIKDFQKDSKAYLDALGNDRIFAFVSKKDKTHSALITEFGNGDLSYTLKDYCK 475
DB 421 KEKEKFEITEIKDFQKDSKAYLDALGNDRIFAFVSKKDKTHSALITEFGNGDLSYTLKDYCK 480

QY 476 KADKALDREKNVTLOGSLKHDGMFVDVSNFYTKYTNASKNPNKGVGVTVNGVSHLEVGFNKV 535
DB 481 KADKALDREKNVTLOGSLKHDGMFVDVSNFYTKYTNASKNPNKGVGVTVNGVSHLEVGFNKV 540

QY 536 AIFNLPDLNLNLAITSFVRNRLNEDKLTGKLSPOEANKLIKDFLSSNKELVGLTKLNFNAV 595
DB 541 AVFNLPDLNLNLAITSFVRNRLNENKLTAKGLSLOEANKLIKDFLSSNKELVGLTKLNFNAV 600

QY 596 ADAKNTGNYDEVKKAQKLESLRKEHLEKEVEKLESKGNKNKMEAKAQAANSOKDEI 655
DB 601 AEAKSTGNYDEVKKAQKLESLRKEHLEKEVEKLESKGNKNKMEAKAQAANSOKDEI 660

QY 656 FALINKEANRDARAAYAQNLKIGIKRELSDKLENVKNLKDFFDKSFPDEPKNGKNDFSKA 715
DB 661 FALINKEANRDARAAYAQNLKIGIKRELSDKLEKISDKLDFSKSFPDEPKNGKNDFSKA 720

QY 716 EETLALKGSVKDLGINTPEWISKVENLNAALNFKNGKNKDFSKYTOAKSDLENSVKOVI 775
DB 721 EETLALKGSVKDLGINTPEWISKVENLNAALNFKNGKNKDFSKYTOAKSDLENSVKOVI 780

QY 776 INOKVTDKVDNLNQAVSAKATGDFSRVBEQALADLKNFSKEOLAQAQKNESLNARKKSE 835
DB 781 INOKVTDKVDNLNQAVSAKMGDFSRVBEQVLADLKNFSKEOLAQAQKNEDFNTGKNSE 840

QY 836 IYQSVKNGVNGTLVGNLSQAATTLKSNFSDIKKELNAKLNNFNNNNNGLKN--EPIY 893
DB 841 LYQSVKNSVNTLVNGLSGIEATLAKNFSDIKKELNEKFNFN--NNNNNGLKNSEPIY 899

QY 894 AKVNKKKAGAAASLEPIYAQVAKVNAKIDRLNOITASGLGVVGOAAGPPLKRHDKVDOL 953
DB 900 AKVNKKKAGAAASLEPIYQVAKVNAKIDRLNOITASGLGVVGOAAGPPLKRHDKVDOL 959

QY 954 SKVGL-----SRNQLAOKIDNLNQAVSEAK 979
DB 960 SKVGLSASPEPIYATIDDLGGPPPLKRHDKVDOLSKVGSRNQELAOKIDNLNQAVSEAK 1019

QY 980 AGFFGNLEQTIDKLKDKSTKHNPMLNLVESAKKVPASLSAKLDNYATNSHIRINSNIKCA 1039
DB 1020 AGFFGNLEQTIDKLKDKSTKHNPMLNLVESAKKVPASLSAKLDNYATNSHIRINSNIKCA 1079

QY 1040 INEKATGMTOKNPEWLKLVNDKIVAHNNVGSVPLSEYDKIGFNQKNMKDYSDSFKESTKL 1099
DB 1080 INEKATGMTOKNPEWLKLVNDKIVAHNNVGSVPLSEYDKIGFNQKNMKDYSDSFKESTKL 1139

QY 1100 NAVKDTSGNSGTQFILTNAFSTASYCYCLARENAEHGKKNVNTKGGFOKS 1147
DB 1140 NNAVXDKIGSFTHFLANAFST-GYYCLARENAEHGKKNVNTKGGFOKS 1186

RESULT 3
CGA2_HELPY
ID CGA2_HELPY STANDARD; PRT: 1182 AA.
AC P5746;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN (120 KDA PROTEIN).
GN CAGA OR CAI.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_taxid=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 53726 / 84-183;
RX MEDLINE=93239281; PubMed=8478069;
RA Tumuru M.K.R., Cover T.L., Blaser M.J.;
RT "Cloning and expression of a high-molecular-mass major antigen of Helicobacter pylori: evidence of linkage to cytotoxin production.";
RL Infect. Immun. 61:1799-1809(1993).
CC !- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L11714; ; NOT_ANNOTATED_CDS.
DR HSP: P02937; 1MLP.
KW Antigen.
FT DOMAIN 878 885 POLY-ASN
SEQUENCE 1182 AA; 131503 MW; C916817E2EE57BB4 CRC64;

Query Match 89.8%; Score 5270.5; DB 1; Length 1182;
Best Local Similarity 88.3%; Pred. No. 2.8e-176;
Matches 1046; Conservative 33; Mismatches 66; Indels 39; Gaps 6;

QY 1 MTNETIDQOQTEAAFPNQOFINNLOVAFKLVNDNAVASYDDPKPIVDKNDNRQAFEG 60
DB 1 MTNETIDQOQTEAAFPNQOFINNLOVAFKLVNDNAVASYDDPKPIVDKNDNRQAFEG 60

QY 61 ISQREYSNKAINKPTKKNOYFSDFINKSNDLINKNDLIDVESSTKSFQFGDQYRIF 120
DB 61 ISQREYSNKAINKPTKKNOYFSDFINKSNDLINKNDLIDVESSTKSFQFGDQYRIF 120

QY 121 TSWVSHONDPSKINTRISIRNPMENIIPPLDDKEAEFLKSAKQSFAGIIGNQIRTDQ 180
DB 121 TSWVSHONDPSKINTRISIRNPMENIIPPIPPDDKEAEFLKSAKQSFAGIIGNQIRTDQ 180

QY 181 KFMGVFDESILKEROEAKNGEPTGGDWLDFLSPFDKQSSDVKEAINOEPVPHVQPD 240
DB 181 KFMGVFDESILKEROEAKNGEPTGGDWLDFLSPFDKQSSDVKEAINOEPVPHVQPD 240

QY 241 ATTTDDIQLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLHNALS 300

```

Db 241 ATSTTHIQGLPPSRDLDERGNFSKFTLGDMEMLDVEGVADMDPNKYKNQLIHNNTLS 300
QY 301 SVLMGSHNGTEPEKVSLLYGGNGGPGCARHDWATVGYKDOQGNVATINVHMKGSGIV 360
Db 301 SVLMGSHDGEPEKVSLLYAGNGGFGAKHDWATVGYKDOQGNVATINVHMKGSGIV 360
QY 361 IAGGEGKINNPSYLYKEDQLTGSRALSQOEIQNKIDPMEFLAONNAKLDNLSEKEK 420
Db 361 IAGGEGKINNPSYLYKEDQLTGSRALSQOEIQNKIDPMEFLAONNAKLDNLSEKEK 420
QY 421 FRTEIKDFQKSKAYLDALGNDRIFAIVSKDKTKHSALITEFNGDLSYTLKDYGKKADKA 480
Db 421 FRNEIKDFQKSKPYLDALGNDRIFAIVSKDKTKHSALITEFNGDLSYTLKDYGKKQIKKA 480
QY 481 LDREKNTVTLQGLSKHDGVNFVDYSNFKYTNASKPNKGVGTNGVSHLEVGNKVAIFNL 540
Db 481 LDREKNTVTLQGLSKHDGVNFVDYSNFKYTNASKPNKGVGTNGVSHLEVGNKVAIFNL 540
QY 541 PDLNLAITSFVRRLNEDLTKTGLSPQANKLIKDFLSNKLKGLTKLNFNAKAVADAKN 600
Db 541 PDLNLAITSFVRRLNEDLTKTGLSPQANKLIKDFLSNKLKGLTKLNFNAKAVAEAKN 600
QY 601 TGNVDEVKKAQDLEKSLKREHLEK-EYEKKLESKSGNKNMEAKAQAANSOKDEIFALI 659
Db 601 TGNVDEVKKAQDLEKSLKREHLEKGDVAKNLESKSGNKNMEAKAQAANSOKDEIFALI 660
QY 660 NKEANDARAIAVAONLKGIGKRELSDKLENVKNLKDQDKSPDEKNGKNKDFSKAEETL 719
Db 661 NKEANDARAIAVAONLKGIGKRELSDKLENVKNLKDQDKSPDGFKNKGNKDFSKAEETL 720
QY 720 KALGSKVDLGINPEWISVENLNAALNEFKNGKNKDFSKVTQAQSDLENSKVDVLIQK 779
Db 721 KALGSKVDLGINPEWISVENLNAALNEFKNGKNKDFSKVTQAQSDQENSIDVLIQK 780
QY 780 VTDKVDNLQAVSVAKATGDFRVEQALADLNKFSKEQLAQAQKNESLNARKKSEIYOS 839
Db 781 ITDKVDNLQAVSVAKIACDFSGVEQALADLNKFSKEQLAQAQKNESFNV-GKSEIYOS 839
QY 840 VKNVNGTLVGNGLSQAEATTLTKNFSKDIKELNKLGNFNNNNNGLKN--EPIYAKVN 897
Db 840 VKNVNGTLVGNGLSGIEATLAKNFSDIKELNKLGNFNFNNNNNGLKNGGEPYIAQVN 899
QY 898 KKKAGQAASLEPIYAAQVAKYNAKIDRLNQIA-SGLGVVQQA----- 939
Db 900 KKKAGQVASEPIYAAQVAKYTKIDQLNQAATSGFGVGQAGPPLKRDHDKVEDLSKVG 959
QY 940 -----AGPPLKRDHDKVDLSKVLGRNOELAQKIDNLNQAVSEAKAGFF 983
Db 960 RSVSPEPIYATIDDLGGSPPLKRDHDKVDLSKVLGRNOELAQKIDNLNQAVSEAKAGFF 1019
QY 984 GNLQETIDKLDKSTKHNPNLWAVESAKKVPASISAKLDNATNSHTRINSNKGAINKEK 1043
Db 1020 GNLQETIDKLDKFTKHNPNLWAVESAKKVPASISAKLDNATNSHTRINSNKGAINKEK 1079
QY 1044 ATCMLTQKNPEWLKLVNDKIVAHNVGSPVLSYDIDKIGFNKKNKMDYSDSFKESTKLNNAV 1103
Db 1080 ATGTERQKNPEWLKLVNDKIVAHNVGSPVLSYDIDKIGFNKKNKMDYSDSFKESTKLNNAV 1139
QY 1104 KDTNSGTOFLNARSTASYCLARENABHGKKNVNTKGGPKS 1147
Db 1140 KDTKSFTOFLANARST-GYSSWARENABHGKKNVNTKGGPKS 1182

```

```

RESULT 4
CAGA_HELPJ
ID CAGA_HELPJ STANDARD: PRT: 1167 AA.
AC Q9ZLT1;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN (120 KDA PROTEIN)
DE (CAG PATHOGENICITY ISLAND PROTEIN 26).

```

```

GN CAGA OR CAI OR CAG26 OR JHP0495.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Gullid B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC - FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001483; AAD06073.1;
KW Antigen; Complete proteome.
FT DOMAIN 246 249 POLY-THR.
FT DOMAIN 882 889 POLY-ASN.
SQ SEQUENCE 1167 AA; 129729 MW; FD5E86881CEBD0F2 CRC64;

```

```

Query Match 86.7%; Score 5086; DB 1; Length 1167;
Best Local Similarity 84.1%; Pred. No. 6.8e-170;
Matches 996; Conservative 60; Mismatches 73; Indels 56; Gaps 4;

QY 1 MTNETIDQOQTEAFANFPOQFINNQLQVAFKVDNVAASVDPOKPIVDKNDNRNQAPEG 60
Db 1 MTNEALNQOQTEAFANFPOQFINNQLQVAFKVDNVAASVDPOKPIVDKNDNRNQAPEG 60
QY 61 ISQLEESYNSKAIKNPTKKNQYFSDFFINKNDLNDLIDVESSTKSFQKGDQRYIF 120
Db 61 ISQLEEFANKAIKNTKKNQYFSSFKSNLDLIDKNDLIDTGSSTKSFQKGTQRYQIF 120
QY 121 TSWVSHQNDPDKINTRIRNFENIIPDLDKEAEPLKSAKQSAFAGIIGNQIRTOQ 180
Db 121 MNWVSHQNDPDKINQKIRGFENIIPDLDKEAEPLKSAKQSAFAGIIGNQIRSDQ 180
QY 181 KPMGVFDESLEKROEAKNGE---PTGGDWLDIFLSFTFDKKQSSDVKEAINEOEPVPHV 236
Db 181 KPMGVFDESLEKROEAKNGEGPDTGGDWLDIFLSFVFNKKQSSDLKETLNQEPVPHV 240
QY 237 QPDIAATTTDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNKYKNQLIHN 296
Db 241 QPDVATTTDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNKYKNQLIHN 300
QY 297 NALSSVLMGSHNGIEPEKVSLLYGGNGGPGCARHDWATVGYKDOQGNVATINVHMKG 356
Db 301 NALSSVLMGSHNGIEPEKVSLLYGGNGGPGCARHDWATVGYKDOQGNVATINVHMKG 360
QY 357 SGLVIAGGEGKINNPSYLYKEDQLTGSRALSQOEIQNKIDPMEFLAONNAKLDNLSEK 416
Db 361 SGLVIAGGEGKINNPSYLYKEDQLTGSRALSQOEIQNKIDPMEFLAONNAKLDNLSEK 420
QY 417 EKEKFRTEIKDFQKSKAYLDALGNDRIFAIVSKDKTKHSALITEFNGDLSYTLKDYGKK 476
Db 421 EKEKFRTEIKDFQKSKAYLDALGNDRIFAIVSKDKTKHSALITEFNGDLSYTLKDYGKK 480
QY 477 ADKALDREKNTVTLQGLSKHDGVNFVDYSNFKYTNASKPNKGVGTNGVSHLEVGNKVA 536
Db 481 ADKALDREKNTVTLQGLSKHDGVNFVDYSNFKYTNASKPNKGVGTNGVSHLEVGNKVA 540

```

```
QY 537 IFNLPDLNLAITSFVRNLEDKLTGKLSPOEANKLIKDFLSSNKELVKGTLNFKAVA 596
DB 541 VFNLPNLAITSFVRQDLEDKLIAGKLSPOEANKLVKDFLSSNKELVKGALNFKAVA 600
QY 597 DAKNTGNYDEVKKAQDKLESLRKRLEKREVEKLESKGNKNKMEAKAQSOKDEIF 656
DB 601 EAKNTGNYDEVKKAQDKLESLRKRLEKREVEKLESKGNKNKMEAKAQSOKDEIF 660
QY 657 ALINKEANRDAIAYAOGLKRELSDKLENVKNKIDKDFDEFFKNGKNDKFSKAE 716
DB 661 ALINKEANRDAIAYAOGLKRELSDKLENVKNKIDKDFDEFFKNGKNDKFSKAE 720
QY 717 ETLKALGSKVDLGINPEWISKVENLAALNEFKNGKNDKFSKVTQAKSDLENSVKDVII 776
DB 721 ETLKALGSKVDLGINPEWISKVENLAALNEFKNGKNDKFSKVTQAKSDLENSIKDVII 780
QY 777 NOKYTDVNDLNQAVSVAKATGDFSRYEQALADLNKFSKEQALAOQAKNESLNARKKSEI 836
DB 781 NOKYTDVNDLNQAVSVAKATGDFSRYEQALADLNKFSKEQALAOQAKNESLNARKKSEI 840
QY 837 YQSVKNGVNGTLVNGLSQAEATTLKSNFSDIKKELNAKLGNNFNNNNGLKNEPIYAKV 896
DB 841 YQSVKNGVNGTLVNGLSQAEATTLKSNFSDIKKELNAKLGNNFNNNNGLKNEPIYAKV 894
QY 897 NKKKAGQAASLEPIYAAQVAKKYNKIDRLNQLIASGLGVGVQAG----- 941
DB 895 -----TEPIYTAQVAKKYNKIDRLNQLIASGLGVGVQAG----- 943
QY 942 -----PLPKRHKVDLISKVGLSRNQLAOKIDNLQAVSEAKAGF 982
DB 944 GLSANHEPIYATIDDLGGPPLKPKRHKVDLISKVGLSRNQLAOKIDNLQAVSEAKASH 1003
QY 983 FGNLEQITDKLSDTKHNPMLWVESAKKVPASLSAKLDNYATNSHINSIKNGAINE 1042
DB 1004 FDNLDQMDLKDSTKKNVNLVYESAKKVPASLSAKLDNYATNSHINSIKNGAINE 1063
QY 1043 KATGMLTKQKPEWKLVDKIVAHNVGSVPLSEYDKIGFNOKNMKDYSDSKFSTKLNNNA 1102
DB 1064 KATGMLTKQKPEWKLVDKIVAHNVGSVPLSEYDKIGFNOKNMKDYSDSKFSTKLNNNA 1123
QY 1103 VKDNTSGFTQTLNASTASYICLARENAEHIKGNVNTKGGFQKS 1147
DB 1124 VKDNTSGFTQTLNASTASYICLARENAEHIKGNVNTKGGFQKS 1167

RESULT 5
USOL_YEAST
ID USOL_YEAST STANDARD; PRT; 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN USOL.
GN USOL OR INT1 OR YD1058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae."
RL J. Cell Biol. 113:245-260(1991).
RN
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RA Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RL
```

```
[3]
RN SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X54378; CAA38253.1; -
CC EMBL; L03188; AAB00143.1; -
CC EMBL; U53668; AAB66659.1; -
CC PIR; A38455; A38455.
CC HSP; P80220; IDIP.
CC SGD; S0002216; USOL.
CC InterPro; IPR002017; Spectrin.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;
```

```
Query Match 5.48; Score 314; DB 1; Length 1790;
Best Local Similarity 18.68; Pred. No. 0.00032;
Matches 233; Conservative 212; Mismatches 475; Indels 330; Gaps 47;

QY 3 NEDSIQQPQTEAFNPOQFINNQLQVAFKVDN-----AVASVDPDQKPI-----VD 48
DB 677 NEDSIQLPELDETGLPKVYFSTY---FIQFNENIYRTALSHDPDEPKISFEEVE 733
QY 49 KNDNRNQAQEGISOLREYSNK-----ATKNPKKNQYFSDFINKNLNDLKNL 99
DB 734 KLQRCQTKLKEITSLQTESTHENLTKLIALNEHKLDEKYQILNSSHSL-KENF 792
QY 100 IDVESSTKSFQFGDQ--RYRIFTSWVSHONDPSKINTSRNFMENLIQ-----PPIL 151
DB 793 SILETELKNVRDSDLEMTQLRDVLETKDKENOTALLEYKSTHKQEDSIKLEKLETL 852
QY 152 DDKEKAE--FLKSQKSFAGIIGNIQNTDQKFMGVDFESLKERQEAENKQEPGGDWLD 209
DB 853 SQKKAEDGINKMKDLPF-----LSREMQAVENCK-----884
QY 210 IFLSFIFDKKQSSDVKEAINOEPVHPQDPIATTTTIDQGLPPEARDLLDERGNFSKFTL 269
DB 885 -----NLQEKDKSNVNHQKESKSLKDAKITEKAINENLEEMKIQCNLSK---934
QY 270 GDMEMLDVEGVADIDPNYKFNOLLIHNNALSSVLMSGSHNGIEPEKVSLLYCGNGPGGARH 329
DB 934 -----NLQEKDKSNVNHQKESKSLKDAKITEKAINENLEEMKIQCNLSK---934
```

Db 935 -EKEHISKELV-----EYK-SRFQSHDNLVAKL-----TEKLKSLAN----- 969

Qy 330 DNNAVYKDDQGNVATINVHMKNGLVYAGGKGINNPSFYLYKEDQLTGSQRLS 389

Db 970 -----NYKDMAENESLIKAV-----EESKNESIQL----- 996

Qy 390 QBEIQNKIDFMEFLAQNNAKLDNLSEKEKEKERTELKDFOKD---SKAYLDALGNDRIF 446

Db 997 -SNLQNKIDSM-----SOEKENFQIERGSTIEKNIEQLKTKTISDLEQTKBEI 1041

Qy 447 VSKKOTKHSALITEFGNGDLSVTLADYKADKALDRKN-----VTLQGSUKHGVVVF 501

Db 1042 ISKSDSSDEYESQIS-----LLKEKETATTANDENVNKISELTKTREELEAEALAYK 1095

Qy 502 DYSNFKYTNASKNPNKGVTVNGVSHL---EYGFNKAIFNLPLDNLNLAITSFVRNLE- 557

Db 1096 NLKNELETLETSEKALKEVENEHLEKEEIQLEKEATEETKQQLNSL-----RANLES 1149

Qy 558 -----DKLTTKGLSPQBAKANKLIDFSSNKELVCK 587

Db 1150 LEKEHEDLAAQLKYEEQIANKEROYNEEISQINDEITSTQOENESIK---KKNDELEGE 1206

Qy 588 T-----LNFENKAVADAKNTGYDEVKKAOKDLE-----KSLRKRREHLEKEVEK 630

Db 1207 VKAMKSTSEEQNLKKSEIDALNL-QIKELKKKNETNEASLLESKVSSESTVKIKELQD 1265

Qy 631 KLESKGNKNKMEAKAQAANSOKDEIFALINKKANRADAIAQAQNLKGIKRELS----- 684

Db 1266 EGNFEKEVESELEDKLKASEDKNSKYLELOKESK-----IKEELDAKTTEL 1312

Qy 685 -----DKLENVKNLKDQPKDFDNKNGKNKDFSKAEETLKALKGVSKDLGINPENWISKE 740

Db 1313 KIQLEKIITNLKAKSESELSRLKKTSEERKNAEQLEKL-----NETQ 1359

Qy 741 NLNAALNEFKGNKDESKVTOAKSDLENSVKOVII-----NOKYTKVDNLNAQVSAK 795

Db 1360 INQAFERKERLLNKGSTITQIYSEKINTLEDELIRLQENELKAKEIDNTRSELEKVS 1419

Qy 796 ATGDFSERVEQALADLNKFS-----KEQLAQAQKNSLNARKKSEIYQSVKNGVNG--- 846

Db 1420 LSND-ELLEKEQNTIKSLQDEILSYKDKITRDNKLLSIEDRNKRD-ESLKEQLRAAQE 1477

Qy 847 --TLVNGLSQAEATTLKSNFSDIK--KELNAKLGNNNNNNGLKNEPIYAKVKNKKAQ 902

Db 1478 SKAKVEELGLKLEES-SKEKAELEKSEMMKLESTIESNETELKSSMETIRKSDEKLE 1536

Qy 903 QA-ASLEEPI-----YAQVAKVNAKIDRLNQLASGLGVVGOAAGFPL---KRHDKVD 951

Db 1537 QSKSABEDIKNLQHEKSDLSRINSEKDIIEELKSLRIEAK-SGSELETVKQELNNAQ 1595

Qy 952 DLSKVGLSRNOELAQKIDNLNQAVSEAKAGFGNLEQ-----TIDKLDSTKHNP 1002

Db 1596 EKIRINAEENTVLKSKLEIDRELKDKQAEIKSQEERELLTSRLKELEQELDSTQ---- 1651

Qy 1003 NLWVESAKVPASLSAKLDNATNSHIRINSNIKANGAINEKATGMTOKNPEWLKLVN-- 1060

Db 1652 ---QKAKSEBERAEYRKF-----QVEKSQDEKAMLETYN---DLVNKE 1693

Qy 1061 ---DKIVAHNVGVPLESDYKIGFNQKNMKDYSDSEKFTKLNNVAKD 1105

Db 1694 QAWKREDTVKTQDSQOEIEKLAKELDLNKAENSKLKEANEDRSEIDD 1743

RESULT 6

RBP2_PLAVB

ID RBP2_PLAVB STANDARD; PRT; 1251 AA.

AC Q00799;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE RETICULOCYTE BINDING PROTEIN 2 (FRAGMENT).

GN RBP2.

OS Plasmodium vivax (strain Belen).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=31273;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92315338; PubMed=1617731;

RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;

RT "A reticulocyte-binding protein complex of Plasmodium vivax

merozoites.";

RL Cell 69:1213-1226(1992).

CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO

HUMAN RETICULOCYTE CELLS.

CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).

CC

CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)

or send an email to license@isb-sib.ch).

CC

DR EMBL; M88098; AAA29744.1; -

KW Malaria; Receptor; Membrane.

FT NON_TER 1

FT NON_TER 1251

SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 5.2%; Score 307; DB 1; Length 1251;

Best Local Similarity 19.0%; Pred. No. 0.00038;

Matches 255; Conservative 221; Mismatches 471; Indels 392; Gaps 65;

Qy 7 DOQPTAEAFNPQPFINLQVAFKVDNAVASYDPDQPIVDKNDNRNRAQFEGISQRLS 66

Db 3 DKEKVDTSFDEKK--KSIEKAYEKMGNTLKE-----LEKMD-DEKNIEKEVEEAAQI 51

Qy 67 EYSNKAINKPTKKNOYSDFFINKSNDLKNKLI-----DYESSTKSFQKFGD 114

Db 52 QY-----KRIFIDHVNLMNDEVEKSVIVMEKTELYKKEIDEIKQKNEVKQDGT 101

Qy 115 QYRIFTSWVSHQNDPSKI-----NTRSTRNFMENIIQPP-----ILDDK 154

Db 102 SNFYTEQYNSATOSKAKIEQFINIATTKGTSDTSQDINELSEIKEEVHKNLQVLQKES 161

Qy 155 EKAFLKSAKSAFAGIIGNQIRT-----DQKFMGVFDESLK----- 191

Db 162 NSMEEMRKQILSMKDLLLNSETIAKELISNTQNALG-FRENATKLNKTDLLQRAA 220

Qy 192 --ERQAEKNGEPTGGDWLIDFLSPFDKQSSDYK--EAINQEPVPHVQPDIAITTTDI 247

Db 221 MIEEAKAHKN-----IDIALE---DAQIDTEVSKIEQINRE-IMNKDEIKSYLSEI 269

Qy 248 QGLPEARDLLDERCNFSKFTLGDMEMLDVEGVADIDNYKFNOLLIIINNALSSVLMGSH 307

Db 270 K-----EYKDKCTTEISNKRQDKKTEFLE----KFKFNEESNSKNVINEINIRSE 320

Qy 308 NGIEPEKVSLLYGGNGPGARHDWNAVTVGYKQDOGNVATINVHMKNGLVGIAGGKG 367

Db 321 QYK-----DIEDAEKQASTKVELFKH-----ETT 346

Qy 368 INNPFSYLYKEDQLTGSQRLSQEIQNKIDFMEFLAQNNAKLDNLSEKEKEKFTKID 427

Db 347 ISN---IFKESEILGVE-TKSQKKINKAEDIMKEIERHSEI-----QTQVKG 390

Qy 428 FQKDSKAY-----LDALGNDR-----IAFVSKKDTKHSALITEFGN---GDLIS 468

Db 391 FOENLNKLNPHNDNAEDLNNDKSTNAKVLITETNESVKHN--LSBITIKQGEKIY 448

Qy 469 T-LKDYGKKADKALDRKNVTLOGSLKHGVMFVDYSNFKYTN-----ASKNPNKGVG-- 520

Db 449 SKADIMQKIKATSENTAEKLE-KVKDDOSNVYVNLQIITERNLIYTERNLNGIDST 507

Qy 521 VTN-----GVSHLEVGFNKVATFNLPDLNLAITSFVRNLEDKLTKGLSPOEANKL 573

Db 508 ITNIEGALKESGNYEIGF-----LEKLEIG-----KNRKLKVDITKSI-----NST 551
Qy 574 IKDFLSSNKLVLGTLNFKAVADAKNTGNYDEVYKKAQKLEKSLRK-REHLEKEVEKKL 632
Db 552 VGNESLFFNFDLNOYDFNKNINDYEN-----KMGIEYNEPEGSINKISELNASENTS 606
Qy 633 ESKSGNKNMEAKAQAANSOKDEIFALINK--EANDARAIAYAOQLKIGIKRELSDKLENV 690
Db 607 DYNIAKTLRLLEA-----OKEV-NNLNKEEBEANKYLRDVKKVESFRFI-FNMKESLDKI 658
Qy 691 NKNLKFDFKSPDEFKNGKNKDFSKAEETLKALKGVSVDLGINPEWISKVENLNAALNEFK 750
Db 659 NEMIK-----KEQLTVNEGHVGNVOLVENIHELVDENLSDILKO-A 699
Qy 751 NGKNKFSKVTOAKSDLENSVKDI-----INOKVTD----- 782
Db 700 TGKNEEQIKITH--STLKNKAKTILGHVDTSAKYVGIKITPELALITELLDGAKLKTAEQ 757
Qy 783 KVDNLNQAV-----SVAKATGDFSRVEQALADLNKFSKEQLAQO-----AOKNESLNARKK 833
Db 758 KFSKNNVVLTEMNKSNKTNELD--VHKNIQDAYKVALEILAHSDIEDTKDKDSSKLIEMG 816
Qy 834 SEIQSV-----KNGVNGTLVGNLSOAEATILSKNFSDIKK-----ELN 873
Db 817 NQIYLVVLINQYKNKISSIKSEAVSVKIGNVSKKHSLSKITCSDKSYDNIIALEKQ 876
Qy 874 AKLGNFN-----NNNN-----GLKN--EPIYAKVNNKKAG-----QAASL 907
Db 877 TELQNLRSFTQETNTNSDSKLEKIKTDFESLKNALKTLEGEVNAKASSDNHEHVQSK 936
Qy 908 EEPYIAQVA--KKYNAKIDRLNQIASGLGVVQQAQFPLKHKVDLDSKVLGRNQEBA 965
Db 937 SEPVPALSETEKEETDIDSLTALDELLKKGRTC--EVSRYKLIKDTVTREISDDTEL- 993
Qy 966 QKIDNLNOAVSEAKAGRFGNLEQITDKL-----KDSKKNPMNLWESAKKVPASL 1016
Db 994 --INTIEKNVAYLAYIKKNYEDTVQDVLTLNEHFNKQVSNHEPTNF--DKSNK-----S 1045
Qy 1017 SAKLDNYATNSH-----IRNSNIKNGAINEKATGMLTOKNPEWLKLVNDKIVAH 1066
Db 1046 SEELTKAVTDSKTIISKLGVIIEVNTENTMTIESSAKEIEALYN-----ELKNKKTSLN 1101
Qy 1067 NV-----GSVPLSEYDKIGFNOKNMKDYDSFKSTKLANNAVKDNTSGTQFLTNAFSTAS 1122
Db 1102 EYQTSNEVKLQE-----MKSNAKDYIDVKIPNTVLDTOK--SNIVTN----- 1143
Qy 1123 YYCLARENAEHGKIKVNTK 1141
Db 1144 -----QHSINNVDK 1153

RESULT 7

YD86_SCHPO STANDARD; PRT; 1957 AA.
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 222.8 KDA PROTEIN CLF3.06C IN CHROMOSOME I.
GN SPAC1F3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN SEQUENCE FROM N.A.
RP STRAIN=972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 270690; CAA94624.1; --
KW Hypothetical protein.
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 5.1%: Score 301; DB 1; Length 1957;
Best Local Similarity 18.3%; Pred. No. 0.001;
Matches 264; Conservative 240; Mismatches 489; Indels 452; Gaps 65;
Qy 2 TNETIDQO--POTEAAPNF--QOFINNLOVAFKVDNNAVASYDDQKPIYDKNDNRNQAFE 59
Db 379 SNRTIHSQTLDAESKLSSEFQENKSLKGS IDEYQNNLSSKDKMKVQVSSQLEEARSLAH 438
Qy 60 GISOLRE-----EYSNKAINKPTKKNQYFDFIN-KSNDLINKONLIDVESTSKSFQKFG 113
Db 439 ATGKLAETINSRDEQNKIKDKFEIQDLRACLSSSNSELKEKSALIDKK----- 488
Qy 114 DQRYRIFTSWYSHONDPSKINTRSIRNPMENIIOPLLDDKEAEFLKS----- 162
Db 489 DOELNNREQIKQKKVSESTQSSLSQ-----LORDIILNKKKHVEYSQLELNELGELQ 543
Qy 163 -----AKQSPAGIIGNQIRTD-----QKFMGVFDE-----SLKREQAEKN 199
Db 544 EISNSEHSSLSQSLTLAAEKEAVATNNELSEKNSLQTLNFAOEKLAQSVOLKENSQN 603
Qy 200 GEPTGGWLDIFLSFIDFKOSSDVKEAINQEPVHPVQDIATTTTIDIQLPPEARLLD 259
Db 604 -----FSSLDTSF--KCLNESHQELENNH-----QTITKQLKDTSSKLQQLQ 644
Qy 260 ERGNF--SKFTLGD-----MEMLDV-----EGVADIDPN----- 286
Db 645 ERANFEQESTLSDENNDLRTKLLKLEESKSLIKQEDVDSLEKNIQTLKEDLRKSEA 704
Qy 287 YKFNOLLTHN-NALSSVLMGSHNGIEPKVSLLYGGNGPGARHDWNAATVGYKDOQGNV 345
Db 705 LRFKLEAKNLREVIDNLKGHETLEA-----RNDLHSSLS----- 741
Qy 346 ATIIINVHMGSLVAGGEGKINNPSFYLYKEDQLTGSQRLSQ-----EEIQN- 395
Db 742 -----DAKNTNAILSELTKSSEDDVKRLTANVETLTQDSKAMKQSFSLVNSYQISNL 795
Qy 396 ----KIDPMEFLAQNNAKLDNLSEKEKEKFEIKDFOKDSKAYLDALGNDRIFAVSKD 451
Db 796 YHELDDHVNMQSQNTLLS-----ESKLTDCENLTQNNMTLI-----DNVQKLMEKH 845
Qy 452 TKHSALITEFG--NGDLSYTLKDYKKADKAL-DREKNVTLQGSGLKHGVMFVDYSNFKY 508
Db 846 VNQESKVELKEVNGKJSLDULNRLSSLNVAISNDQILTQLAELSK-----NIDSLEQ 899
Qy 509 TNASKNPN-KGVGVTVNGVSH-----LEVGFNKVA-----IFN 539
Db 900 ESAQLNSGLKSLEAKQLLHTENEELHRLDLKLTGKLTIEESKSSDLGKLTARQEEISN 959
Qy 540 LPDLN---NLAITSFVRNLEDKITTKGLSPQEAANKLKD-----FLS 579
Db 960 LKEENMSQSAITS-VKSKLDETLS-----KSKLEADIEHLKNKVSEVEVERNALLA 1011
Qy 580 SNKELVGKTLNFKAVAD-----AKN-----TGNYDEV-----KKAQKOLEK 616
Db 1012 SNEKMLDDKNNGENIASLQTEIEKKRAENDDLQSKLSVSEYENLLILSSQNTKSLD 1071
Qy 617 SLRREHLEKEVEKKLESKKNKME-----AKAANSQ-KDEIFALINKEANRDAR 668
Db 1072 KTNQLKYEKNVQKLLDEKQDNQVNELELTSKYKGLGENAQIKDELLALRKKSKQHD 1131
Qy 669 AIAYAQNLIKGIKRELSDKLENVNNKLDKDFDKSDFEFKNGKNKDFSKAETLALKGSVKD 728

Db 215 BEKQYISVLQTVSLLKQRLNRNGPMNVDLKPLPQLPQAEVFTKEENPESDGPVVED 274
Qy 207 WLDIFLSIFD---KKQSDVKEINQPPVHPQDPDIATTTDQGLPPEARDLLDERGN 263
Db 275 GTSVKLTETLOQVKRQENLLKRC--KETIQSHREQCTLLTSEKAL-----QEQLDER-- 326
Qy 264 FSKFTLGDMEMLDVEGVADIDPNYK---FNOLLTHNNAALSSVLMGSHNGIEPEKVSLLYG 320
Db 327 -----LQELKIKDLHMAEKTLLITQRLDAKNLIEQ-----LEQDKGMV1-- 366
Qy 321 GNGGPGARHDWNAVGYKDQGNVATINVHMKGSL-----VIAGEKKGINN 370
Db 367 ---AETKROMHETLEKKEEIAQLRSRIKQMTTQGEELREQEKESRAAFEELEKALST 422
Qy 371 P-----SYLYKEDQLTQSQRALSOBEI-----QNKIDMEFLAQN--AKLD 411
Db 423 AQKTEEARRLKLAEMDEQIKTSEBETISLQELSRVQEVVDVNMKSSSEQIAKLQ 482
Qy 412 NLSEKE-----KEKPRTEIKDFORDSKAYLDALGNDRITAFVSKKDKTHS----- 455
Db 483 KLHEKELARKEQELTKLQTREREFQOMKVALEKSOSEYLKISOEKBQESLAELEL 542
Qy 456 ---ALITFPGNGDLSYTLKDYKGRADKALDRKNVTLQGSCLKHDGVMFVDYSNFKYTNAS 512
Db 543 QKKAILESEN-----KLRLDQEAETV--RTRILELESSLE-----KSLOEN 583
Qy 513 KPNKGVGTNGVSHLEVGFNKVAIFNLPLNLNLAITSFVR-----RNLEDKLTTK 563
Db 584 KNQSKDLAV-----HLEAEKKN-----HNKEITVMVEKHKTELESLEKHOQDALWTE 629
Qy 564 GLS-----POEANKLIKD-----FLSNKELVGKTLNPNKAVADAKNTG 602
Db 630 KLQVLKQOYQTEMKLEKREKCEQETLLKKEIIFQAHIEEMNEKTLK-----KLDVKQT- 684
Qy 603 NYDEVKRAQKLSLRKREHLEVEVEKLESKGNKNKMEAK--AOANSOKDEIFALIN 660
Db 685 ---ELESLSSELSVLKARHKLEELSVLKQOTDKMKQLEAKMDEQKNHHQOQVDSIIK 741
Qy 661 KEANRDARAIYAQNLKGIKRELSKDLNVNKNLKDDEKS-----700
Db 742 E-----HEVSIQRTKALKADQINQLELLKRLKDKHLKEHQAHVENLEADIKRSEG 791
Qy 701 -----FDEFKNGKNKDFKAETTLKALGSKVKKDLGI-----731
Db 792 ELQOASAKLDVFSQSAETHQTKAYEQALQLOOKLQDLLETERILLTKQVAEVEAQKDK 851
Qy 732 -----NPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLEN 769
Db 852 VCTELDAHKIQVDLMQOLEKQNSEMEQKVSKLTQVYESKLEDGKQEQTKILLVEKEN 911
Qy 770 SV-----KDV-IINQKVTDKVDN-----LNOAVSVAKATGDFSRVEQALADLKNFSKEQ 817
Db 912 MILQREGOKREIEILTQKLSAKEDSIHLNEEVET-----KFRNQEKKMEKVQKAKEM 966
Qy 818 LAQAQKNESLN--ARKSEIYQSVKNGVNTLVNGLSQAEBATLTKNFSIDIKKELNAK 875
Db 967 --QETLKKLLDQEAELKEL-----ENTAL--ELSQKEKQFNAK 1002
Qy 876 LGFNENNNNGLKNEPIYAKVNNKAGQAASLEPIYAVAKKVNAKIDRLNQAISGLGV 935
Db 1003 MLEMAQANSAGISDAVSRLTNQKE--QIESLTVHRRELNDVISTWEKKLUNQQA----- 1055
Qy 936 VQGAAGPPLKRHKVDLLSKYGLSRNQ-----ELAQKI-----DNLNOAVSEAKAGPFG 984
Db 1056 -----EELQETHEIQLOEKEQEAELVAELKQILLFGCEKEMKNKEITLWKEGV- 1102
Qy 985 NLEQIDKLKSTHNPMLNVESAKKVPASLSAKLDNYAT--NSHTRININIKNGAINE 1042
Db 1103 KQDTTLNLEQBLQKSAH--VNSLAQDETTLKAKHLEKLEVDLKNLSKENTFLQEQVL 1160
Qy 1043 KATGMLTOKNPEWLKLVNDKIVAHNVGSPVL--SEYDKIGFNQKNKMDYSDSFK-----FS 1096
Db 1161 K---MLAEDRRKVSLETSKLTDEEPQSLKSSHEK---SNKLEDSLEFKKLSSELA 1214

Qy 1097 TKLNNAVKDTNS---GFTQFLTNAFSTASYCYCLAR-ENAEH 1133
Db 1215 IQLDICCKKTEALAEKATNELINISSKTNAILSRISHCOH 1255

RESULT 9

CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CENTROMERIC PROTEIN E (CENP-E PROTEIN).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.F., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
interactions with the kinetochore proteins CENP-F and hBUBR1.";
J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).
CC
CC EMBL: Z15005; CAA78727.1; -
DR PIR: S28261; S28261.
DR HSP: P03069; 1211.
DR MIN: 117143; -
DR InterPro: IPR001752; kinesin.
DR Pfam: PF00225; kinesin; 1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SM00129; KISC; 1.
DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE: PS00067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335 MECHANOCHEMICAL (MOTOR).
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

COILED COIL (POTENTIAL).

FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
 FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
 FT CONFLICT 301 301 R -> A (IN REF. 1).
 SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 4.8%; Score 284; DB 1; Length 1875;
 Best Local Similarity 18.6%; Pred. No. 0.0037;
 Matches 253; Conservative 225; Mismatches 493; Indels 392; Gaps 57;

QY 3 NETIDQOPOTAAFPQOFI-----NNLOVAFKVDNNAVASYDDPKPIVDKNDK 52
 DB 375 NENLSAKSSDFELKOLIKERTKEHLQNOIETFI VELEHKV-----PIIN----- 422

QY 53 DNQAF-EGISQLREEYSNKAI---KNPTKKNOYFSDFINKSNLDINKNDLIDVESSTKS 108
 DB 423 ---SFKERTDMLNENLNAALLLEHTSNEKNNAKVKELNAKQKLVCEEN-DLQTLTKQ 476

QY 109 FOKGDO-RYRIFTSWSH-----QNDPSKINTRSIRNF 141
 DB 477 RDLCRQIYLLITNSVNSDSKGPLRKEEIOFIQIMOEDDSTITESDSQKVVTIRLVEF 536

QY 142 MENIOPPILDDKEAEFLKSAKOSFAGILIGNOLRTDQKMGVDESLSKEROE----- 196
 DB 537 -KNIIQ---LOEK-NAELLKVR-NLADKLESKEKSKQSLQKIESETVNNAKALITLK 590

QY 197 -EKNGEPTGDMWLDIFLFDKQSSDVKEAINQEPVPHVOPDIATTTTIOGLPPPEAR 255
 DB 591 SEK-----MDLESRIEOLKELEBELKTSVPNEDASYSNVITIKQLTETKRDLESQVQ 641

QY 256 DLLDERGNSFTGLDMEMLDVEGVADIDPNYKFNOLLIHNNALSSVLMSGHNGIEPEKV 315
 DB 642 DLQTRISQITRESTENMSLLNKE-IDQLYDSKSDISIKLGKESKRILAEERFKLLSNLT 700

QY 316 SLLYGNGCGCARHDWNAVTVGKDOGNVATIIIVHKNKSGVLVAGGKGIN---NPS 372
 DB 701 DLTRAENDQLKRFDYLONTILK-QDSKTHETLNEYVSCSKLSIVETEL-LNLKEBQK 757

QY 373 FYLKEDQLTGSRALSQEE-----IQNKIDFMEFLAQ 405
 DB 758 LRVLHKLKQLNKLSPKSDSLRIMVTOLOTLOKEREDELLEETRKSCQKIDLE----- 813

QY 406 NNAKLDNLSEKEKEFRTE--IKFOKDSKA-----YLDALGNDRIFAFVSKDKTKHSA 456
 DB 814 -----DALSELKETSQKDHKIOLEEDNNNIEWYQNKIEALKDYESVITSVDKOTD 868

QY 457 LITEFGNDLSYTKDYCK--KADKALDRKNV-----TLOGSLKHGVMFVD-YSN 505
 DB 869 I-----EKLOYKVKSLSEKEIEEDKIRLHTYVMDDETINDSLRKEKSKINLTDAVSQ 922

QY 506 FK-YTNASKNPKNGVGTNGVSHLEVGFNKVA--IFNLPDLNLAITSFVRNLEDK--- 559
 DB 923 IKEYKDYETTSQSLOQTN--SKLDESFKDTNQIKNLT-----EKTSLEDKISL 971

QY 560 -----LTTKGLSQEAN--KLIDFLSSNKLKELVKTILNFNKAVADAKN----- 600
 DB 972 LKEQMFNLNLDLQKGMEXEKADFKKRISILQNNNKEVEAVKSEYKSLKIQNDLQD 1031

QY 601 -----TGNVDEVKKAQKDLKSLRK-----REHLEKEVEKKLES 634
 DB 1032 QTIYANTQANYEQELQKHADVSKTISELRQLHTYKQVKTILNSRQLENALKENKS 1091

QY 635 KSGNKNK-MAKAQANSQKD-----EIFALINKAN----- 664
 DB 1092 WSSQKESLLEQLDLSNRIEDLSQNKLLYDQIQIYTAADKEVNNTNGPGLNLTITLR 1151

QY 665 -----RDARAIYAQNKLKIG-----QNVAPIESELTAUKYSMOE 680

DB 1152 RERDILTQVTAERDAKMLKQKISLMDVDELQDARTKLDNSRVEKENHSSIIQOHHIME 1211

QY 681 -----RE-----LSDKLENNKLNKDFDKSFDKGNKGNKDFSKAEETLKALKGSKVD 728
 DB 1212 KLNQNLNRESNITURNELENNNNKKKLEQSELDK-----QNVAPIESELTAUKYSMOE 1267

Query Match 4.8%; Score 279; DB 1; Length 1679;
 Best Local Similarity 18.5%; Pred. No. 0.0049;
 Matches 258; Conservative 222; Mismatches 490; Indels 426; Gaps 56;

QY 729 LGINPEWTSKVENLNAALNEFKNGKNKDFSKVTOAKSOLNSVKNVDIINQKVTDKVDML- 787
 DB 1268 -----KEQELKLAAKEEVHRNKKRS-QDILEKHEGLSSS-----DYKLESEIENLK 1312

QY 788 ---NQAVSVAKATGDFSRVEQALADLKNFKEQLAQAQAKNESLNARKKSEIYQSVKNG 843
 DB 1313 ELENKERQGAEEKFNRLRRQ-----AERLKTSLKSLQDLSLEQVNSLRDA 1360

QY 844 VNGTLVGNLGSQAEATTLTKNFSDIKKELNAKLNAGNFNNNNNGLKNEPIYAKVNNKKAGQ 903
 DB 1361 KN--VLENSLSEANAR-----IEELONAKVAGNNOLEAIRKLOEDAESKRE---L 1407

QY 904 AASLEEPI--YAQVAKVNAKIDRLNQIASGLVGVVQAAGFPLKRHDKVDLDSKVLGRN 961
 DB 1408 QAKLEESTSYESTINGLNEETITLKEIE-----KQRIQQOQLQATSANEQ 1454

QY 962 QELAOKIDNLNOAVSEAKAGFGNLEQTI-DKLKDKSTK--HNPMLWVESAKKVPAS--- 1015
 DB 1455 NDLNIVESMKKSFEEDIKFIKEKTOEVNKEILEAQERLNOPSINNEIEIKKKWSEHE 1514

QY 1016 --LSAKL--DNYATNSHIRINSNIKNGAINKATGMLTKPKNPEWKLKVDNKIVAHNVGSV 1071
 DB 1515 QEVQKIREAEALKKRLPT-----EKKINKIIEKKKEELEKEFEKVEERIKSME 1567

QY 1072 PLSEYD-----KIGFNQKNMKDYSDSPKFKSTKLNNAVVD 1105
 DB 1568 QSGETIDVVRKLEAKVQEKOLEN-----EYNKKLQEEELKD 1605

RESULT 11
 YIO9_YEAST STANDARD; PRT; 1679 AA.
 AC P40457;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE HYPOTHETICAL 195.1 KDA PROTEIN IN DNA43-UB11 INTERGENIC REGION.
 GN YIL149C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN=S288C / AB972;
 RA Barrill B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Consey T., Dear S., Devlin K., Fraser A.,
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z38059; CAA86129.1;
 DR PIR; S48385; S48385.
 DR SGD; S0001411; MLP2.
 KW Hypothetical protein.
 SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;

QY 19 QOFINNLOVAELKVDNAVASYDDOKPIVDKNDNRDNRQAFEGIS-----QLREE----- 67
 Db 38 EBEVTKNLVVDIEIKSYQYRSRISKULQLLDESSEKQNTAKEELNGKLDOLNEERSYRRE 97
 QY 68 -----YSNKAIK--NPTKK-----NOYFSDFTKNSDLIN----- 95
 Db 98 IDALKQLHVSHEANREVDKRVKEEYDIWQSRQOGNDSLNDLKNKLLRRKLMEME 157
 QY 96 -----KONLID-----VESSTKSFQK-----FG 113
 Db 158 NILQCKSNALSQLKYDTVSQEKELMLQSKLIEEKLSSFSKTKLTTEBVTKSSHVENLE 217
 QY 114 DORYRI-----FTSWSHONDPKINTRISINFMENIIPILDDK-----EKAELFK 161
 Db 218 EKLYQMOSNYESVFTYNKFLNQN---KQLSQSVE---EKVLEMKNLKDTASVEKAEEFSK 271
 QY 162 S-AKOSFAGIIGNOIRTDQKFMGVFDESLEKQEAENG--EPTGGDWLDFISFIQDK 218
 Db 272 EMTLOKNNDLRSOLTSLK-----DCSLRAIEKNDNDSNCRNPEHTVIDELIDTKURL 326
 QY 219 KOSSD-----VKEAINQEPVPHVQPDIAATTTDQGLPPEARDLLDERGNFSK 266
 Db 327 EKSNECORLQNVMDCTKEBEATWTSVAVPTVGKLFSDIKVL---KROLIKERN--QK 381
 QY 267 FTLG-----DMEMDVEGVADIDPNYKFNOLLIHNNALSSVLMGSHNGIEPEKVSLL 318
 Db 382 FOLQNOLEDIFILEHKTPELISFKERTKSLHEHLKRSLETTELTVSLTKRQEREITSRL 441
 QY 319 YGNGGPGARHDWNAVTVYKQOQGNVATIIINVHMKNGSGLVIAGGEKGINPSFYLYKE 378
 Db 442 QKING-----CEANIISLVKQRLDLARQVKLLNLNTSAI-----QETASP-----LSQ 484
 QY 379 DOLTSORALSQOEIQONIKDFMEFLAQNNAKLDNLSEKEKEKFRPEIKDFQDKSKAYLDA 438
 Db 485 DELISLRKILESNNVENDSQAIIITERLVEFSNVNELQERN-----VEL 529
 QY 439 LGNDRIAFVSKDKTHSALITFEFGNDLSYTLKDYGGKADKALDREKNVTLOGSLKHGVD 498
 Db 530 LNCIRI-----LADKLENYEGKQDKTLQKVENOTIKEA--KDAI 566
 QY 499 MFVDYSNFKYTN-----ASKNPNGVGTNGVSHLEVGFNK-----VAIF 538
 Db 567 TELENINAKMETRINILLRERDSYKLLASTEENK--ANTNSVTSMEAREKRIELELAEL 624
 QY 539 NLPDLNLNLAITSFVR-----NLEDKLTTKGLSPQEAANKL-----IKDFL 578
 Db 625 SSTKVENSAILONLRKELLYKKSOCKKKTTLEDENFKGLAKEKERMLEAIDHLKAE 684
 QY 579 SSNKELVGKTLNFKAVADAKNTGNYDEVKKAQDKLESLRK-----REHL----- 624
 Db 685 EKQKSWPSPYIHVEKERASTELSSOSRIKISLEVEISK-LKKETASFTPTRESLTDPEQ 743
 QY 625 ----EKEVEKKLESK--SGNKNMEAKAQAQNSQDEIFALINK-----EA 663
 Db 744 CKEKELQMRKESEIHNENKMDFFSKQGOYKAKIKELNENLERLSQLSQIKQIES 803
 QY 664 NRDAR--AIAYAQN-----LKGIKRELSDK---LNVNKNLKDFOKSPDEFK----- 705
 Db 804 IRSCKDSQLKWAQNTIDDTMKMKSLLTSLNSKETTIEKLSSEIENLDKELRTKQYQV 863
 QY 706 NGKNKDFSKAETLK-----ALKGSVKOLGINPENWISKVENLNAALNEFKN----- 751
 Db 864 LDONSDASTLEPTLRKELEQIQVOLKDKANSQIQAYEEIISNNEN---ALIELKRELAKTK 920
 QY 752 -----GKNKDFSKV-----TOAKSDLENSVKDVI 775
 Db 921 ENYDAKIELEKKEKWARDEDLSRLGELGETRALQPKLKEGALHFVQOSEKLRNEVERKI- 979
 QY 776 INOKVTDKVDNLNOAVSVAKATGDFSRVEQALADLNKFSK--EQLAQQAQKNESLNARKK 833
 Db 980 --QKMIKIEKMTIVQLCKKK-EMSOYQSTMKENKNDLSSELVIRLEKDAADCAQELTKTK 1036
 QY 834 SEIQSVKNGVNGVLVGNGLSQAEATTLTKNFSKIDKELNAKLNFNNNNNNGLKNEPIY 893

Db 1037 SLSYA-----QDLLOKHHERKWEKADYEREL---ISNIEOTESURVENSVL 1082
 QY 894 AKVKKKA--GOAASLE-EPIYAQVAKKVNNAKIDRLNQIASGLGVVGOAAGPPLKRHDVK 950
 Db 1083 EKVDVDTAANNQDGHKLKLVLSFSLNHRNSLEYTKLTCTCKRELAFVKQ-----KN 1132
 QY 951 DLSKV--GLSRNQLAQIDNLNOAVSEAKAGFFGNLEQITDKLKOSTKHPMNLWVES 1008
 Db 1133 DSLEKTINDLQRTQTLSEKQCSAVI-----IDEPKDTIK----- 1168
 QY 1009 AKVPASLSAKLDNVTNHSIRINSNIKANGAINKATGMLTKQKNPEWKLKLVNDKIVAHNV 1068
 Db 1169 -----EVTQVNILKENNAILOKSLKNVTEKKNREIYKQLNDR--QEEI 1208
 QY 1069 GSVF--LSEYDKIGFNQNMKDYSDSPKFTK---LNNVAKDTNSGFTOFLTNAFSTA 1121
 Db 1209 SRLQDRLIQTQEQVINSKILVYESEMEQCKQRYQDLSQOQKDAQKKDIEKLTNEISDL 1268
 QY 1122 SYCLARENAEHGKN 1137
 Db 1269 KGKLSSAENANADLEN 1284

RESULT 12
 RBPI_PLAVB
 ID RBPI_PLAVB STANDARD; PRT; 2869 AA.
 AC 000798;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
 GN RBPI.
 OS Plasmodium vivax (strain Belem).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 merozoites";
 RL Cell 69:1213-1226(1992).
 CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 HUMAN RETICULOCYTE CELLS.
 CC -!- SUBUNIT: HOMODIMER (POTENTIAL).
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL; M88097; AAA29743.1; -;
 DR HSP; P36956; IAW9.
 DR Malaria; Receptor; Signal; Transmembrane.
 KW SIGNAL
 FT CHAIN 1 17 POTENTIAL.
 FT DOMAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
 FT TRANSMEM 18 2807 EXTRACELLULAR.
 FT TRANSMEM 2808 2826 POTENTIAL.
 FT DOMAIN 2827 2869 CYTOPLASMIC.
 FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
 SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Query Match 4.7% Score 276.5; DB 1; Length 2869;
 Best Local Similarity 19.1%; Pred. No. 0.011;
 Matches 265; Conservative 196; Mismatches 451; Indels 479; Gaps 63;

Db 858 ND-----MRTKFKFNQINKLK-----NDLOMESKKKFFLEEK-----NOKTVNEL 898
Qy 574 --IKDFLSSNNKELVGKTLNFNAKAVADAKNTGNYDEVKKAQKDL--EKSLLRKRHLE----- 625
Db 899 ENTODLLNOEKE-----NLRKNESLNVRKTS--ETLQKQFDLVSERKLEVAQN 953
Qy 626 -KEYEKLKESGNNKMEAKAQ--ANSQKDEIFALINKENRDRATAYQANLKGIRKL 683
Db 954 LEEAHQIQGLQETIRREATLEKLHKKNNELIKQIS--DLNCD-----ISKQSSQSLIKES 1009
Qy 684 SDKLENNKLNKDPDKDFDKNGKKNDFKAEETLKALKGVKDLGINPEWISKVENLN 743
Db 1010 KKLLENKIKRLKDVINKSEKEIKSFNDKLSSEE-----DLDI--KLVTLEKNCN 1057
Qy 744 AALNEFKN-----GKNKDFSKVTQA--KSDLENSVKDVIINQKVTDKVDN----- 786
Db 1058 IAMSRLQSLVTSNDSLRKSNENFKKKAALNNLNKKNKESELL-----KMKKIDNHKKELAT 1114
Qy 787 ----LNOAVSV--AKATGDFSRVEQALADLK-----NFSKEQLAQAAQKN-----E 826
Db 1115 FSKORDDAVSEHGHTAIETALQETRIQLTEYKSNYQKIKEYSNFQRETKEQOKRKNLSVE 1174
Qy 827 SLN-----ARKKSIYOSVKNVNGTLVGNGL--SOAEATTLKSNFSD-----IK 869
Db 1175 SLNDSKIKELEARLSQEI--SLNQYLKRIKSGNSVETNISSTRSTSYSDPLDKEDIK 1232
Qy 870 KELNAKLGNFNNNNNKNEPIYAKYKNNK-----AGQAAASLEEPIYQAQVAKK 918
Db 1233 KYDQLQA--FTEIPRN--LENE-----IEEKNLISRLRFTETRLASSFED-----QK 1278
Qy 919 VNAKIDRLNQIASLG---VVGQAAGPLKR--HDKYDDLSKVGLSRNOELAQKID----- 969
Db 1279 IKAQMKKLKLIQMDPSIPLDSILNEPLDNCPOKESDINKLMLEVY--LKRQLDIETRA 1337
Qy 970 --NLNOAVSEAKAGF-----PGNLEPOTIKLKOSTKKNPM-----NLW 1005
Db 1338 HYDENALISALHFKRIQGESSLSSSDIYKLPFEASEERVKSLDKLTKMPLKDRTNLP 1397
Qy 1006 VESAKKVPASLSA-----KLDNYATNSHIRINSNIKNGAINEKATGMUTOKNPEWLK 1057
Db 1398 VGDIKKNRDSISKYEERYKYLENYKLEITL-----NESNGKLSQILDLRQSKSKEAL-- 1452
Qy 1058 LVNDKIYAHNVGVSPLSE--YDKIGFNQNMKDYSDSPKSTFKLNAVYKDTNSGTQBLTN 1116
Db 1453 -----LSEQLDRL-----QKDLSETERQKELLSTIKQKQKQPCMCMD 1491
Qy 1117 AFST---ASYCYCLARENAEHGKNNV 1138
Db 1492 LQGNELRLREHIALKQAEEDVRNN 1516

RESULT 14

ANT1_ONCVO STANDARD; PRT; 2022 AA.
AC P21249;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MAJOR ANTIGEN.
GN OVT1.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95287898; PubMed=7770081;
RA Tritreaprapab S., Richie T.L., Tuan R.S., Shepley K.J., Dinman J.D.,
RA Neubert T.A., Scott A.L.
RT "Molecular cloning of a gene expressed during early embryonic
development in Onchocerca volvulus.";
RL Mol. Biochem. Parasitol. 69:161-171(1995).

RN SEQUENCE OF 733-866 FROM N.A.
RP MEDLINE=89127417; PubMed=2464764;
RA Luciferson J.E., Duke B.O.L., Moser D., Zeng W., Erundu N.E.,
RA Doneson R., Renz A., Karam M., Flores G.Z.;
RT "Construction of Onchocerca volvulus cDNA libraries and partial
characterization of the cDNA for a major antigen.";
RL Mol. Biochem. Parasitol. 31:241-250(1988).
CC -!- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
with the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U12681; AAA80009.1; -;
DR EMBL; J03995; AAA29412.1; -;
DR PIR; A54513; A54513.
DR HSP; P02633; 3ICB.
KW Antigen; Coiled coil.
FT DOMAIN 74 120 COILED COIL (POTENTIAL).
FT DOMAIN 151 251 COILED COIL (POTENTIAL).
FT DOMAIN 327 384 COILED COIL (POTENTIAL).
FT DOMAIN 417 1879 COILED COIL (POTENTIAL).
SQ SEQUENCE 2022 AA; 237341 MW; B7132AACF1520317 CRC64;

Query Match 4.6%; Score 272.5; DB 1; Length 2022;

Best local Similarity 19.0%; Pred. No. 0.01;
Matches 225; Conservative 213; Mismatches 416; Indels 329; Gaps 57;

Qy 1 MTNETI-----DQOPTEAAFPNQFINNQVAFK-----VDNAVASVYDPQKPIV 47
Db 285 ITQDTVLIEAVKRPHESSQQQTAPVIGP--ELITELRAEADAGLHDELMRYESAKRII 343
Qy 48 DKNDNRQAFEGISQLEESYNNK--AIKNPTKKNQFSDFINKSNLDINKNLDIVESS 105
Db 344 ELEARD-----DESHNKLVAESDLKRT--DRLAESQNALRK--LYDM--- 383
Qy 106 TKSQKQFGDQRYRFTSWSHONDPKINRSIRNF-----ENITQPTLDDKEKAEF 159
Db 384 TYSYEINAEKRETSPTSPTKGVPPPEVV--RSRYVLNSRANDNNVLQKLNKAEVQISE 441
Qy 160 LKSAQKQFAGIIGNQIRTDQKPMGVFDESUKERQAEKNGEPTGGDWLDFLSFIDKK 219
Db 442 LTKNDSLEI-----RRLEKRIAEANRTTHQRELDLDAKHVVKD--LEDRLKSLQEK 495
Qy 220 QSDVKEAINQEPVPHVQPDIAITTTTIDQGLPPPEAR-----LLDERGNFSKFTLGDM--- 272
Db 496 ASDSARHLEDEIRKMRQEPNSTLLDVERAAEDADERIRKIDEE--TKIRISELTNR 552
Qy 273 -EMLDVEG-----VADIDPNYK--FNOLLIHNALSSVLMGSH---NGIEPE 313
Db 553 IEMLEENKRLKDNMGKNRIQDIEKEYNTIIRKLEEKONALKNLENTRQLRVNELEE 612
Qy 314 KVSLLYGGNGPGARHDM-----NATVGVKDOQGNVATTINV-----HMKNG 356
Db 613 R-----TRFDTMTSEFDNLRNTNDSANKNTVAIELTVKQKEQDEIKQDK 660
Qy 357 SGLVIAGEKGINNPSFYLYKEDQLTQSQALSQEEIQNKIDFMFLAQNNAKLDNIS-- 414
Db 661 LAKELADLENKLN-----ETKMRGDAEKLNRHL--DEID--NFKQINNEYITEVTII 710
Qy 415 EKEKEKFEITEIKDPKQSKAYLDALGNDRIAFVSKKDKTKHSALETEFGNGDLSTLKDYG 474
Db 711 RRQNDDEFTQM---KTQAKLSMKNSLIA--AKKETEK---LSEMNN----- 750
Qy 475 KKADKALDRKNVTLOGLSKHDGVFVDYSNFKYTNASKNPKNGVGTNGVSHLEVGFNK 534
Db 751 -----RLQODKNLDLIGAKQKQDTELNL-----LTEKIRKVEIEFER 786

Qy 747 NEFKNGKNKDFSKVTOAKSDLENSVKDVI INOKVTDKVDNLNOA-----VSVAKATG 798
Db 642 KKYLSGNAKDLDN-TNLMEILNDKIS--VLQOQITDVKDELVDSEEREEREAIVAGOKLSA 698
Qy 799 DF---SRVEQALADLKNFSEKQALQAOAKNESLNA-----RKKSEIYO-----SVK 841
Db 699 SFELMSNEKQAL-----ELKYSSLKNELINAQNLDRREELSSELSKKLFEERKIR 749
Qy 842 NGVNGTLVGNCLSOAEATTLTKNFSIDIKKELNAKLGNNNNNNNNGLKNEPIYAKVNKKKA 901
Db 750 SGSNDIDIEKN-----KEINVLNSELADLQAIIRHLESKMDKLMELDKLVHHLN--RG 796
Qy 902 GOAASLEE-----PIYAOVAKKVNAKIDRLNGLASGLGVVGQAA 940
Db 797 LEEANIEENAVKKRLLCLLMGCDYSSVILQIVSQIEHFVNOQIQITIRSLKQEL----- 849
Qy 941 GFPLKRHD-----KVDDL-----SRQELAQKIDNLNQAVSEAK 979
Db 850 -----RHDFVQFSGKKEQELSRSFEKFGTETKHDILAQRRNVSEKMDLENAQK-- 902
Qy 980 AGFTGN-----LEOTIDKL-----KDKTK 998
Db 903 --FFSPDRKNGYLPSEHTSKIEYLEKTIEDLKLALQDELKRNLLMDDISSYNKQTTK 960
Qy 999 HNPMLAVESAKKVPASLSAKLDNYATNSHIRINSNIKNGAINEK----- 1043
Db 961 LOEKIKWLERERSI---LIDELESYRSNQF-----NYONNLVQDKNELEERKEIQKELE 1012
Qy 1044 -----ATCMLTQKNPEWKLKLVNDKIVAHNVGSVPLSEYDKIGFNOKNKNKDYSD 1091
Db 1013 VYNNHFMKQAEMLTSNVTDESQMLKTLREALQSKTNNIDHLSTI--LERNRKEYKSLLD 1070
Qy 1092 SF-----KFSTKLNNAVKTNSGFTQFLTNAFSTASYCYCLARENAEHGK 1136
Db 1071 DYNQLRARYKNLQSNTPQSTQSGOVESEIKGLSKLTKYLOSKCRREHSR 1120

Search completed: February 13, 2002, 14:13:29
Job time: 668 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2002, 14:01:01 ; Search time 118.81 Seconds
(without alignments)
1412.123 Million cell updates/sec

Title: US-09-360-685A-5
Perfect score: 5867
Sequence: 1 MTNETIDQOQTAAAFNQQ.....RENAHGINKVNTKGFQKS 1147

Scoring table: BLOSUM62
Gapop 10.0 , Gapext.0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5859	99.9	1147	2	Q9F220 helicobacte
2	5550	94.6	1179	2	Q9L5X9 helicobacte
3	5448	92.9	1179	2	Q9F222 helicobacte
4	5403	92.1	1247	2	Q9F115 helicobacte
5	5394	91.9	1247	2	Q9F223 helicobacte
6	5352	91.2	1165	2	Q9F221 helicobacte
7	5308	90.5	1247	2	Q07910 helicobacte
8	5263	89.7	1183	2	O86064 helicobacte
9	4741	80.8	1163	2	Q9F230 helicobacte
10	4720	80.4	1171	2	Q9F229 helicobacte
11	4703.5	80.2	1170	2	Q9F218 helicobacte
12	4700	80.1	1173	2	Q9F116 helicobacte
13	4690.5	79.9	1176	2	Q9F231 helicobacte
14	4690.5	79.9	1176	2	Q9F228 helicobacte
15	4689	79.9	1177	2	Q9F225 helicobacte
16	4679	79.8	1173	2	Q9KHU3 helicobacte
17	4678.5	79.7	1230	2	Q9F232 helicobacte
18	4677.5	79.7	1172	2	Q9F227 helicobacte
19	4664.5	79.5	1186	2	Q9F224 helicobacte

20	4658.5	79.4	1172	2	Q9F217 helicobacte
21	4640	79.1	1227	2	Q9F219 helicobacte
22	4635.5	79.0	1172	2	Q9F226 helicobacte
23	1745.5	29.8	441	2	Q9L7K3 helicobacte
24	1718	29.3	408	2	Q9L7K4 helicobacte
25	1695.5	28.9	445	2	Q9L7K5 helicobacte
26	838	14.3	187	2	Q9X5H4 helicobacte
27	764	13.0	149	2	Q9RLX1 helicobacte
28	763	13.0	149	2	Q9RLX1 helicobacte
29	762	13.0	149	2	Q9RLX7 helicobacte
30	760	13.0	149	2	Q9RLX3 helicobacte
31	759	12.9	149	2	Q9RLX4 helicobacte
32	758	12.9	149	2	Q9RLX0 helicobacte
33	758	12.9	149	2	Q9RLX6 helicobacte
34	757	12.9	149	2	Q9RLX5 helicobacte
35	757	12.9	149	2	Q9RLX8 helicobacte
36	756	12.9	149	2	Q9RLX2 helicobacte
37	754	12.9	149	2	Q9RLX5 helicobacte
38	751	12.8	149	2	Q9RLX9 helicobacte
39	749	12.8	149	2	Q9RLX7 helicobacte
40	749	12.8	149	2	Q9RLX6 helicobacte
41	749	12.8	149	2	Q9RLX3 helicobacte
42	748	12.7	149	2	Q9RLX2 helicobacte
43	746	12.7	149	2	Q9RLX8 helicobacte
44	731	12.5	149	2	Q9RLX9 helicobacte
45	718	12.2	149	2	Q9RLX0 helicobacte

ALIGNMENTS

RESULT 1

Q9F220 ID Q9F220 PRELIMINARY; PRT; 1147 AA.
AC Q9F220;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCTC11637;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H., Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from Japanese and non-Japanese isolates.";
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB015416; BAB20926.1;
SQ SEQUENCE 1147 AA; 128003 MW; D740A7275A9B141D CRC64;

Query Match 99.9%; Score 5859; DB 2; Length 1147;
Best Local Similarity 99.9%; Pred. No. 9,9e-238;
Matches 1146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MTNETIDQOQTAAAFNQQFINNLOVAFKVDNAVAYDDPQKPIVDKNDNRDNRQAFEG	60
Db	1	MTNETIDQOQTAAAFNQQFINNLOVAFKVDNAVAYDDPQKPIVDKNDNRDNRQAFEG	60
Qy	61	ISOLREESNKAIKNPTKKNQYFSDFINKSNLIDVESSTKSFQKFGDQRYRIF	120
Db	61	ISOLREESNKAIKNPTKKNQYFSDFINKSNLIDVESSTKSFQKFGDQRYRIF	120
Qy	121	TSWSHQNDPSKINRSIRNFNMENIIQPPILDDKKAELKSAKOSFAGIIIGNIQTQD	180
Db	121	TSWSHQNDPSKINRSIRNFNMENIIQPPILDDKKAELKSAKOSFAGIIIGNIQTQD	180
Qy	181	KPMGVFDSLSKQERQAEKNGEPTGGDWLIDFLSFIDFKQSSDVKEAINQEPVPHVQDDI	240

Db 181 KFMGVFDESLEKQAEKNGEPTGGDWLDIFLSFDFKQSSDVKEAINQEPVPHVQDI 240
Qy 241 ATTTDIOGLPEARDLDDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
Db 241 ATTTDIOGLPEARDLDDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
Qy 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDWNATVGYKDDQGNVATTINVHMKNGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDWNATVGYKDDQGNVATTINVHMKNGLV 360
Qy 361 IAGEKGINNPSFYLYKEDQLTGSGRALSQEEIQNKIDFMEFLAQNNAKLDNLSEKEK 420
Db 361 IAGEKGINNPSFYLYKEDQLTGSGRALSQEEIQNKIDFMEFLAQNNAKLDNLSEKEK 420
Qy 421 FRTEKDFQKDSKAYLDALGNDRITAFVSKDTKHSALITEFGNGDLSYTLKDYKKADKA 480
Db 421 FRTEKDFQKDSKAYLDALGNDRITAFVSKDTKHSALITEFGNGDLSYTLKDYKKADKA 480
Qy 481 LDREKNVTLOGSLKHGDMVDFVSNFYKTNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMVDFVSNFYKTNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
Qy 541 PDNLNLAITSFVRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTLPNFNKAVADAKN 600
Db 541 PDNLNLAITSFVRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTLPNFNKAVADAKN 600
Qy 601 TGNVDEVKKAQKDLKSLRKEHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Db 601 TGNVDEVKKAQKDLKSLRKEHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Qy 661 KEANRDARAIYAQNLIKRELSKLENVKNLKDPSKDFEFKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIYAQNLIKRELSKLENVKNLKDPSKDFEFKNGKNKDFSKAEETLK 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Qy 781 TDKVDNLNQAVSVAKATGDFSRVQALADLNKFSKEQLAQQAQAKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNQAVSVAKATGDFSRVQALADLNKFSKEQLAQQAQAKNESLNARKKSEIYQSV 840
Qy 841 KNGVGTGLVNGLSQAEATTLSKNFSDIKELNAKLNKFNNNNNGLKNEPIYAKVKKK 900
Db 841 KNGVGTGLVNGLSQAEATTLSKNFSDIKELNAKLNKFNNNNNGLKNEPIYAKVKKK 900
Qy 901 AGAASLEPIYAQVAKVKNNAKIDRLNQIASGLGVGQAGFPPLKRHDQVDDLSKVGLSR 960
Db 901 AGAASLEPIYAQVAKVKNNAKIDRLNQIASGLGVGQAGFPPLKRHDQVDDLSKVGLSR 960
Qy 961 NOELAQKIDNLNQAVSEAKAGFFGNLEQTIDKLDKSTKHNPMLNWLWESAKKVPASLSAKL 1020
Db 961 NOELAQKIDNLNQAVSEAKAGFFGNLEQTIDKLDKSTKHNPMLNWLWESAKKVPASLSAKL 1020
Qy 1021 DNYATNSHIRINSINKGAINKEKATGMLTQKNPEWLKLVNDKIVAHNVGVSPLSEYDKIG 1080
Db 1021 DNYATNSHIRINSINKGAINKEKATGMLTQKNPEWLKLVNDKIVAHNVGVSPLSEYDKIG 1080
Qy 1081 FNQNMKDYSDSFSTKLNNAVKDNTSGFTQFLTNAPSTASYCYCLARENAEHGKKNVT 1140
Db 1081 FNQNMKDYSDSFSTKLNNAVKDNTSGFTQFLTNAPSTASYCYCLARENAEHGKKNVT 1140
Qy 1141 KGGFQKS 1147
Db 1141 KGGFQKS 1147

RESULT 2
Q9L5X9 PRELIMINARY; PRT: 1179 AA.
ID Q9L5X9;
AC Q9L5X9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CYTOTOXIN-ASSOCIATED PROTEIN CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN (1) SEQUENCE FROM N.A.
RP STRAIN=SS1;
RC Zhang J.2.;
RA "cagA gene of Helicobacter pylori (SS1 strain).";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN (2) SEQUENCE FROM N.A.
RP STRAIN=SS1;
RC Zhou J.C.; Zhang J.2.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF247651; AAF63759.1;
DR EMBL: AF247651; AAF63759.1;
SQ SEQUENCE 1179 AA; 131297 MW; 7A9112F58B749787 CRC64;

Query Match 94.6%; Score 5550; DB 2; Length 1179;
Best Local Similarity 92.3%; Pred. No. 8.e-225;
Matches 1090; Conservative 19; Mismatches 36; Indels 36; Gaps 3;

Qy 1 MTNETIDQOPQTEAAFNPOQFINNLOVAFKVDNNAVASYDPQKPIVDKNDNRQAE 60
Db 1 MTNETINQOPQTEAAFNPOQFINNLOVAFKVDNNAVASYDPQKPIVDKNDNRQAE 60
Qy 61 ISQLEESNKAINKPTKKNOVFSDFINKSNDLKNLIDVESSTKSFQFGDQRYIF 120
Db 61 ISQLEESNKAINKPTKKNOVFSDFINKSNDLKNLIDVESSTKSFQFGDQRYIF 120
Qy 121 TSWVSHQNDPSKINTSRINFEMENIQQPIIDDKKEAEFLKSQKSAFAGIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTSRINFEMENIQQPIIDDKKEAEFLKSQKSAFAGIIGNQIRTDQ 180
Qy 181 KFMGVFDESLEKQAEKNGEPTGGDWLDIFLSFDFKQSSDVKEAINQEPVPHVQDI 240
Db 181 KFMGVFDESLEKQAEKNGEPTGGDWLDIFLSFDFKQSSDVKEAINQEPVPHVQDI 240
Qy 241 ATTTDIOGLPEARDLDDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
Db 241 ATTTDIOGLPEARDLDDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
Qy 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDWNATVGYKDDQGNVATTINVHMKNGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDWNATVGYKDDQGNVATTINVHMKNGLV 360
Qy 361 IAGEKGINNPSFYLYKEDQLTGSGRALSQEEIQNKIDFMEFLAQNNAKLDNLSEKEK 420
Db 361 IAGEKGINNPSFYLYKEDQLTGSGRALSQEEIQNKIDFMEFLAQNNAKLDNLSEKEK 420
Qy 421 FRTEKDFQKDSKAYLDALGNDRITAFVSKDTKHSALITEFGNGDLSYTLKDYKKADKA 480
Db 421 FRTEKDFQKDSKAYLDALGNDRITAFVSKDTKHSALITEFGNGDLSYTLKDYKKADKA 480
Qy 481 LDREKNVTLOGSLKHGDMVDFVSNFYKTNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMVDFVSNFYKTNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
Qy 541 PDNLNLAITSFVRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTLPNFNKAVADAKN 600
Db 541 PDNLNLAITSFVRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTLPNFNKAVADAKN 600
Qy 601 TGNVDEVKKAQKDLKSLRKEHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Db 601 TGNVDEVKKAQKDLKSLRKEHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Qy 661 KEANRDARAIYAQNLIKRELSKLENVKNLKDPSKDFEFKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIYAQNLIKRELSKLENVKNLKDPSKDFEFKNGKNKDFSKAEETLK 720

```
Db 661 KEANRARAIAATNLRGVRELSKLENVKNLKDPSFDEPKNGKNKDFSKAETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFRNGKNKDFSKVTQAKSDLENSVKDVIINOKV 780
Db 721 TLKGSVKDLGINPEWISKVENLNAALNEFRNGKNKDFSKVTQAKSDLENSVKDVIINOKI 780
QY 781 TDKVDNLNOAVSAKATGDFSRVEQALADLNKFSKEQLAQQAOKNESINARKKSEIYQSV 840
Db 781 TDKVDNLNOAVSAKATGDFSRVEQALADLNKFSKEQLAQQAOKNEFNGKNSALYQSV 840
QY 841 KNGVNGTLVGLNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Db 841 KNGVNGTLVGLNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
QY 901 AGQAASLEPIYAOVAKKVNKIDRLNQIASGLGVQQAAGFPPLKRHDKV 951
Db 901 AGQVSPSEPIYAOVAKKVNKIDRLNQIASGLGVQQAAGFPPLKRHDKV 959
QY 952 -----DLSKVGLSRQELAQKIDNLNOAVSEAKAGFFGNL 986
Db 960 SPEPIYATIDDLGPPFLKHKAKVEDLSKVGLSRQELTOKIDNLNOAVSEAKAGFFGNL 1019
QY 987 EOTIDKLDSTKHPMNLWVESAKKVPASLSAKLDNYATNSHIRINSNIKNGAINERKATG 1046
Db 1020 EOTIDKLDSTKHPMNLWVESAKKVPASLSAKLDNYATNSHTRINSNIQNGAINERKATG 1079
QY 1047 MLTQKNPEWLKLVNDKIVAHNVGSPVSEYDKIGFNOKNMKDYSDSPKSTKLNNAVKDT 1106
Db 1080 MLTQKNPEWLKLVNDKIVAHNVGSPVSEYDKIGFNOKNMKDYSDSPKSTKLNNAVKDV 1139
QY 1107 NSGFTQFLTNAFSTASYCYCLARENAEHGKKNVNTKGGFOKS 1147
Db 1140 RSGETQFLTNAFST-GYICLAENAEHGIENVNNTKGGFOKS 1179

RESULT 3
Q9F222 ID Q9F222 PRELIMINARY; PRT; 1179 AA.
AC Q9F222;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC43579;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB015414; BAB20924.1; -
SQ SEQUENCE 1179 AA; 131686 MW; 7D0ADC34E644634 CRC64;

Query Match 92.9%; Score 5448; DB 2; Length 1179;
Best Local Similarity 90.9%; Pred. No. 1.6e-220;
Matches 1073; Conservative 23; Mismatches 50; Indels 34; Gaps 2;

QY 1 MTNETIDQQPQTEAFAFPQFQINNQLQVAFKVDNAVASYDPDQKPIVDKDRDRNRAFEQ 60
Db 1 MTNETINQQPQTEAFAFPQFQINNQLQVAFKVDNAVASYDPDQKPIVDKDRDRNRAFEQ 60
QY 61 ISQLEEYSNKAIKNPTKKQYFSDFINKSNLDINKNLDIVESTSKSFQKFGDQRYRIF 120
Db 61 ISQLEEYSNKAIKNPTKKQYFSDFINKSNLDINKNLDIVESTSKSFQKFGDQRYRIF 120
QY 121 TSWVSHONDPSKINTRSRNFMENIQQPIPDDEKAEFLSAKQSFAGIIVGNQIRTDQ 180
```

```
Q9RF15
ID Q9RF15 PRELIMINARY; PRT; 1247 AA.
AC Q9RF15;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11637;
RA Ito Y., Azuma T.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF202973; AAF17598.1; -.
SQ SEQUENCE 1247 AA; 139042 MW; 28DEC5888F9D041B CRC64;

Query Match 92.1%; Score 5403; DB 2; Length 1247;
Best Local Similarity 85.5%; Pred. No. 1.3e-218;
Matches 1067; Conservative 35; Mismatches 44; Indels 102; Gaps 2;

Qy 1 MTNETIDQPOTEAAFPNQQFNNLQVAFKVDNAVASYDPDQKPIVDKDRDRNRQAFEG 60
Db 1 MTNETIDQPOTEAAFPNQQFNNLQVAFKVDNAVASYDPDQKPIVDKDRDRNRQAFEG 60

Qy 61 ISQREEYSNKAINKPTKKNOYFSDFINKSNDLNKNDLIDVESSTKSFQFGQORYIF 120
Db 61 ISQREEYSNKAINKPTKKNOYFSDFINKSNDLNKNDLIDVESSTKSFQFGQORYIF 120

Qy 121 TSWYSHQNDPSKINTSRIRNMENTIOPILDDKEAEFLKSAKSPAGIIIGNOIRTDQ 180
Db 121 TSWYSHQNDPSKINTSRIRNMENTIOPILDDKEAEFLKSAKSPAGIIIGNOIRTDQ 180

Qy 181 KFMGVDESLKERQEAENKGEPTGGDMLDIFLSIFDKQSSDYKKEAINQEPVPHVQPD 240
Db 181 KFMGVDEFLKERQEAENKGEPTGGDMLDIFLSFVENKEQSSDYKKEAINQEPVPHVQPD 240

Qy 241 ATTTTDTQGLPEARDLLDERGNSKFTLGDMEMLDVEGVADIPNFKFNOILLHNALS 300
Db 241 ATTTTTHQGLPEARDLLDERGNSKFTLGDMEMLDVEGVADIPNFKFNOILLHNALS 300

Qy 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHWNATVGVKQOQGNVATTINVHMKNGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNTVGVYKNOQGNVATTINVHMKNGLV 360

Qy 361 IAGGEGINNPSFYLYKEDQITQSORALSQBEIQNKIDFMFLAQNNAKLDNLSEKEK 420
Db 361 IAGGEGINNPSFYLYKEDQITQSORALSQBEIRNKIDFMFLAQNNAKLDNLSEKEK 420

Qy 421 FRTIKDFQKDSKAYLDALGNDRITAFVSKDKTKHSALITEFGNGDLSYTLKDYCKKADKA 480
Db 421 FQNEIEDFQKDSKAYLDALGNDRITAFVSKDKTKHSALITEFGKGDLSYTLKDYCKKADRA 480

Qy 481 LDREKNVTLOGSLKHGDMFVDYSNFYKTNASKNPKNKGVGTNGVSHLEVGFKNKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMFVDYSNFYKTNASKNPKNKGVGTNGVSHLDAGFSKVAIFNL 540

Qy 541 PDLNNLAITSFVRNLEDKLTKTKSQEANKLIKDFLSSNKNELVGKTLNFKNAVADAKN 600
Db 541 PDLNNLAITSFVRNLENKLVTEGLSLQEANKLIKDFLSSNKNELVGRALNFKNKAVADAKN 600

Qy 601 TGNVDEVYKKAQKDLKSLRKEHLEKEVEKKLESKSNKNKMEAKAANSQKDIIFALIN 660
Db 601 TGNVDEVYKKAQKDLKSLRKEHLEKEVEKKLESKSNKNKMEAKAANSQKDKIFALIN 660

Qy 661 KEANRDRATAYAQNLKIGKIRELSDKLENNVKNLKDQKSFDEPKNGKNKDFSKAETLK 720
Db 661 KEANRDRATAYSQNLKIGKIRELSDKLEKINKDLKDFSKSFDEPKNGKNKDFSKAETLK 720
```

```
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKFSKVTOAKSDLENSKVDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKFSKVTOAKSDLENSKVDVIINQKI 780

Qy 781 TDKVDNLNQAQVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNQAQVSMARATGDFSRVEQALADLKNFSKEQLAQQAQTKNESEFNVGKKSEIYQSV 840

Qy 841 KNGVNGTLVGNGLSQAEATLTSKNFSDIKKELNAKILGNFNNNNNNGLKNEPIYAKVNNKK 900
Db 841 KNGVNGTLVGNGLSGIEATLAKNFSDIKKELNEKFKNPNNNNNNNGLENEPIYAKVNNKK 900

Qy 901 AGQAASLEPIYIAQVAKVNAKIDRLNQITASGLGVVQA----- 939
Db 901 TGQVASPEPIYIAQVAKVNAKIDRLNQAAASGLGGVQAGFPLKRHDKVDLDSKYGRSVS 960

Qy 940 ----- 939
Db 940 ----- 939

Qy 961 PEPIYATIDDLGGPPPLKRHDKVDDL SKVGRSVSPEIYATIDDLGGPPPLKRHDKVDL 1020
Db 961 -----AGFPLKRHDKVDDL SKVGLSRNOELAQKIDNLNQAVSEAK 979

Qy 1021 SKVGRSVSPEIYATIDDLGGPPPLKRHDKVDDL SKVGLSRNOELAQKIDNLNQAVSEAK 1080
Db 1021 SKVGRSVSPEIYATIDDLGGPPPLKRHDKVDDL SKVGLSRNOELAQKIDNLNQAVSEAK 1080

Qy 980 AGFEGNLEQITDKLKDSTKHNPMLNWSAKVPASLSAKLDNYATNSHIRINSNIKNGA 1039
Db 1081 AGFESNLEQITDKLKDSTKYNSVNLWESAKVPASLSAKLDNYATNSHTRINSIQNGA 1140

Qy 1040 INEKATGMLTOKNPWLKLVNDKIVAHNVGSLSEYDKIGNQKNMKDYSDSFSTKL 1099
Db 1141 INEKATGMLTOKNPWLKLVNDKIVAHNVGSLSEYDKIGNQKNMKDYSDSFSTKL 1200

Qy 1100 NNAVDTNSGTQFTLTNAFSTASYCLARENAEHGKIKVNTKGGFQKS 1147
Db 1201 NNAVDTNSGTQFTLTNAFST--GYISLARENAEHGKIKVNTKGGFQKS 1247

RESULT 5
Q9F223 PRELIMINARY; PRT; 1247 AA.
ID Q9F223
AC Q9F223;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC43526;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RL J Gastroenterol. 35:890-897(2000).
DR EMBL; AB015413; BAB20923.1; -.
SQ SEQUENCE 1247 AA; 139064 MW; 121E68825E5C9228 CRC64;

Query Match 91.9%; Score 5394; DB 2; Length 1247;
Best Local Similarity 85.3%; Pred. No. 3.2e-218;
Matches 1065; Conservative 35; Mismatches 46; Indels 102; Gaps 2;

Qy 1 MTNETIDQPOTEAAFPNQQFNNLQVAFKVDNAVASYDPDQKPIVDKDRDRNRQAFEG 60
Db 1 MTNETIDQPOTEAAFPNQQFNNLQVAFKVDNAVASYDPDQKPIVDKDRDRNRQAFEG 60

Qy 61 ISQREEYSNKAINKPTKKNOYFSDFINKSNDLNKNDLIDVESSTKSFQFGQORYIF 120
Db 61 ISQREEYSNKAINKPTKKNOYFSDFINKSNDLNKNDLIDVESSTKSFQFGQORYIF 120
```



```
QY 121 TSWVSHONDPSKINTRSIRNFEMENIIQPPILDDKEAEFLKSAKQSPAGIIGNQIRTDQ 180
Db 121 TSWVSHONDPSKINTRSIRNFEMENIIQPPILDDKEAEFLKSAKQSPAGIIGNQIRTDQ 180
QY 181 KFMGVDESILKEROEAKNGEPTGGDWLDFLSFIPDKKQSSDVKEAINQEPVPHVQPD 240
Db 181 KFMGVDEFELKEROEAKNGEPTGGDWLDFLSFVFNKEQSSDVKEAINQEPVPHVQPD 240
QY 241 ATTTTIDQGLPPARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
Db 241 ATTTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKDOQGNVATIIIVHMKNGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNATVGYKNOQGDVNATIIIVHMKNGSLV 360
QY 361 IAGEGKGINNPSFYLKEDQLTGSORALSQEIEONKIDPMEFLAONNAKLDNLSEKKEK 420
Db 361 IAGEGKGINNPSFYLKEDQLTGSORALSQEIEIRNKIDPMEFLAONNAKLDNLSEKKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAIVSKKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 FONEIEDFOKDSKAYLDALGNDRIFAIVSKKDPKHSALITEFGKGDLSYTLKDYGKKADRA 480
QY 481 LDREKNVTLOGSLKHGDMVFDVSNFKYTNASKNPKNKGVTNGVSHLEVGFNKVAIFNL 540
Db 481 LDREKNVTLOGNKLKHDGVMFVNTSNFKYTNASKSPDKGVGTNGVSHLDAGFSKVAIFNL 540
QY 541 PDLNNLAITSFVRNLEDKLTTKGLSQEANKLIKDFLSSNKELVCKTLNFKNAVADAKN 600
Db 541 PDLNNLAITSFVRNLENKLTITGLSLQEAANKLIKDFLSSNKELVGLKALNFNAVADAKN 600
QY 601 TGNVDEVKAQKDLERSLRKREHLEKEVEKKLESKSGKNKMEAKAQAQNSOKDEIFALIN 660
Db 601 TGNVDEVKAQKDFERSLRKREHLEKEVEKKLESKSGKNKMEAKAQAQNSOKDKIFALIN 660
QY 661 KEANRARAIAQAQNLKIGIKRELSDKLENVKNLKDPSDFEFKNGKNDKFSKAEETLK 720
Db 661 KEANRARAIAQAQNLKIGIKRELSDKLEKINKDLKDPFSKDFEFKNGKNDKFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLAALNEFKNGKNKOPSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLAALNEFKNGKNKOPSKVTOAKSDLENSVKDVIINQKI 780
QY 781 TDKVDNLNQAVSAKATGDFSRVEQALADLNKFSKEQALQAKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNQAVSAKATGDFSRVEQALADLNKFSKEQALQAKNESFNVKKSEIYQSV 840
QY 841 KNGVNGTLVGNGLSQABATTLKSNFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNNKK 900
Db 841 KNGVNGTLVGNGLSGTEATALAKNFSDIKKELNEKFKFNNNNNNNGLNEPIYAKVNNKK 900
QY 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVQCA----- 939
Db 901 TGQVASPEPIYAQVAKVNAKIDRLNQASGLGVQAGFPLKRHKDKVDLDSKVGSRVS 960
QY 940 ----- 939
Db 940 ----- 939
QY 961 PEPIYATIDDLGGPFPLKRHKVDLDSKVGSRVSPEPIYATIDDLGGPFPLKRHKVDL 1020
QY 940 -----AGFPLKRHKVDLDSKVGSLSRNQELAQKIDNLNQAVSEAK 979
Db 1021 SKVGSRVSPEPIYATIDDLGGPFPLKRHKVDLDSKVGSLSRNQELAQKIDNLQAVSEAK 1080
QY 980 AGFEFNGLEQIDKLKOSTKINPNLWVESAKYPASLSAKLDNYATNSHTRINSINIGA 1039
Db 1081 AGFEFNGLEQIDKLKOSTKYNSNLWVESAKYPASLSAKLDNYATNSHTRINSINIOGA 1140
QY 1040 INEKATGMLTOKNPEWKLKLVNDKIVAHNVGSVPLSEYDKTGFNQKNMKDYSDFKFSKTL 1099
Db 1141 INEKATGMLTOKNPEWKLKLVNDKIVAHNVGSVPLSEYDKTGFNQKNMKDYSDFKFSKTL 1200
```

```
QY 1100 NNAVKDNTNSGFTQFLTNAPSTASYCYCLARENAEHGKIKNVNTKGGFOKS 1147
Db 1201 NNAVKDVKSSFTOFLANAST-GYISLARENAEHGKIKNVNTKGGFOKS 1247
RESULT 6
Q9F221 PRELIMINARY; PRT: 1165 AA.
AC Q9F221;
DT 01-MAR-2001 (Tremblurel. 16, Created)
DT 01-MAR-2001 (Tremblurel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblurel. 16, Last annotation update)
DE .CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
ON NCBI_TaxID=210;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC49503;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RT J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB015415; BAB20925.1; -.
SQ SEQUENCE 1165 AA; 130140 MW; 99C316921E649C02 CRC64;

Query Match 91.2%; Score 5352; DB 2; Length 1165;
Best Local Similarity 90.4%; Pred. No. 1.7e-216;
Matches 1057; Conservative 28; Mismatches 58; Indels 26; Gaps 7;

QY 1 MTNETIDQQPTEAAFPQPFINNLOVAFLKVDNAVASYPDQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETAQQQPTEAAFPQPFINNLOVAFLKVDNAVASYPDQKPIVDKNDNRDNRQAFEG 60
QY 61 ISQLEEYSNKAINKPTKKNOYFSDFINKSNDLINKNDLIDVESSTKSFKQFGQRYRIF 120
Db 61 ISQLEEYSNKAINKPTKKNOYFSDFINKSNDLINKNDLIDVESSTKSFKQFGQRYRIF 120
QY 121 TSWVSHONDPSKINTRSIRNFEMENIIQPPILDDKEAEFLKSAKQSPAGIIGNQIRTDQ 180
Db 121 TSWVSHONDPSKINTRSIRNFEMENIIQPPILDDKEAEFLKSAKQSPAGIIGNQIRTDQ 180
QY 181 KFMGVDESILKEROEAKNGEPTGGDWLDFLSFIPDKKQSSDVKEAINQEPVPHVQPD 240
Db 181 KFMGVDESILKEROEAKNGSTGGDWLDFLSFIPDKKQSSDVKEAINQEPVPHVQPD 240
QY 241 ATTTTIDQGLPPARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
Db 241 ATTTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKDOQGNVATIIIVHMKNGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNATVGYKDOQGNVATIIIVHMKNGSLV 360
QY 361 IAGEGKGINNPSFYLKEDQLTGSORALSQEIEONKIDPMEFLAONNAKLDNLSEKKEK 420
Db 361 IAGEGKGINNPSFYLKEDQLTGSORALSQEIEIRNKIDPMEFLAONNAKLDNLSEKKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAIVSKKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 FRTEIKDFQKDSKAYLDALGNDRIFAIVSKKDTKHSALITEFGKGDLSYTLKDYGKKADKA 480
QY 481 LDREKNVTLOGSLKHGDMVFDVSNFKYTNASKNPKNKGVTNGVSHLEVGFNKVAIFNL 540
Db 481 LDREKNVTLOGNKLKHDGVMFVDVSNFKYTNASKNPKNKGVTNGVSHLDAGFSKVAIFNF 540
QY 541 PDLNNLAITSFVRNLEDKLTTKGLSQEANKLIKDFLSSNKELVCKTLNFKNAVADAKN 600
Db 541 PDLNNLAITSFVRNLEDKLAAKGLSQEANKLIKDFLSSNKELVGLKALNFNAVADAKN 600
```

```

QY 601 TGNVDEVKKAQKDLKSLRKHLEKEVEKKLEKSGKNGKMKMEAKAQAQNSOKDEIFALIN 660
DB 601 TGNVDEVKKAQKDLKSLRKHLEKEVEKKLEKSGKNGKMKMEAKAQAQNSOKDEIFALIN 660
QY 661 KEANRDARAIYAONLKGIGIKRELSDKLENVKNLKDPSKDFEPKNGKNDKDFSKAEETLK 720
DB 661 KEANREARAIYAONLKGIGIKRELSDKLENVKNLKDPSKDFEPKNGKNDKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKDFSKVTOAKSDLENSKVDIINOKV 780
DB 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKDFSKVTOAKSDLENSKVDIINOKV 780
QY 781 TDKVDNLNQAVSVAKATGDFSRVQALADLNKFSKEQLAQQAQKNSNARKKSEIYQSV 840
DB 781 TDKVDNLNQAVSVAKATGDFSRVQALADLNKFSKEQLAQQAQKNSNARKKSEIYQSV 840
QY 841 KNGVNGTLVGNGLSOAEATTLKSNFSDIKKELNAKLNKFNNNNNGLKN--EPIYAKVKN 898
DB 841 KNSVNTLVGNGLSGIEATLAKNFSDIKELNEKFNKFN--NNNNNGFNKFNSTPIYAKVKN 899
QY 899 KKQAQASLEPIYAQVAKV-----NAKIDRLNQAISGL-----GVVGOA 939
DB 900 KKTGOVASPEPIYAQVKNKKKAGQAGPPLKRHDKVDLDSKV--GLSASPEPIYATIDDL 957
QY 940 AG-FPLKRHKVDLDSKVLGRNQLAQKIDNLAQVSEAKAGFFGNLEQITDKLKDSTK 998
DB 958 GGPEPLKRHKVDLDSKVLGRNQLAQKIDNLAQVSEAKAGFFGNLEQITDKLKDFTK 1017
QY 999 HNPMLWVESAKVPASLSAKLDNYATNSHTRINSNTKNGAINKATCMLTOKNPEWLK 1058
DB 1018 HNPMLWVESAKVPTLSAKLDNYATNSHTRINSNTKNGAINKATCMLTOKNPEWLK 1077
QY 1059 VNDKIVAHNVGSVPLSEYDKIGFQKNNKMDYSDSKFESTKLNNNAVKDNTSGFTQFLTNAF 1118
DB 1078 VNDKIVAHNVGSVPLSEYDKIGFQKNNKMDYSDSKFESTKLNNNAVKDNTSGFTQFLTNAF 1137
QY 1119 STASYICLARENAEHGKNNVNTKGGFQKS 1147
DB 1138 ST-GYYCLARENAEHGKNNVNTKGGFQKS 1165

RESULT 7
ID 007910 PRELIMINARY; PRT: 1247 AA.
AC 007910;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CAGA (CYTOTOXIN ASSOCIATED PROTEIN A).
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43526;
RA Maeda S., Kanai F., Ogura K., Yoshida H., Ikenoue T., Kawabe T.,
RA Shiratori Y., Omata M.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC43526;
RA Maeda S., Kanai F., Ogura K., Yoshida H., Ikenoue T., Kawabe T.,
RA Shiratori Y., Omata M.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 335-415 FROM N.A.
RC STRAIN=NCTC11637;
RA Auerman M., Azuma T., Berg D.E., Ito Y., Morelli G., Pan Z.J.,
RA Suerbaum S., Thompson S., van der Ende A., van Doorn L.J.;
RT "Recombination and clonal groupings within Helicobacter pylori from

```

```

RT different geographic regions.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001357; AAB58747.1; -
DR EMBL; AB003397; BAA19853.1; -
DR EMBL; AJ239729; CAB37821.1; -
SQ SEQUENCE 1247 AA; 138819 MW; 172C76AE9F9BD29E CRC64;

Query Match          90.5%; Score 5308; DB 2; Length 1247;
Best Local Similarity 84.1%; Pred. No. 1.3e-214;
Matches 1051; Conservative 41; Mismatches 53; Indels 104; Gaps 3;

QY 1 MTNETIDQOPQTEAAFNPPQFINNLOVAFLVDNNAVSYDPQKPIVDKNDNRNDRQAPEG 60
DB 1 MTNETIAQOPQTEAAFNPPQFINNLOVAFLVDNNAVSYDPQKPIVDKNDNRNDRQAPEG 60
QY 61 ISOLREESNAIKNPTTKNQYFSDFINKSNLDINKNDLIDVESSTKSFQFGDORYRIF 120
DB 61 ISOLREESNAIKNPTTKNQYFSDFINKSNLDINKNDLIDVESSTKSFQFGDORYRIF 120
QY 121 TSWYSHONDPSKINTSRIRNFEMENIQPPILDDKEKAEFLKSAKOSFAGIIIGNOIRTDQ 180
DB 121 TSWYSHONDPSKINTSRIRNFEMENIQPPTPDDKERAEFVSAKOSFAGIIIGNOIRTDQ 180
QY 181 KFMGVFDESLEKQEAKEKNGEPTGGDWLDIFLSFIFDKKQSSDVKEAINQBPVPHVQPD 240
DB 181 KFMGVFDEFLKQEAKEKNGEPTGGDWLDIFLSFVFENKEQSSDVKEAINQBPVPHVQPD 240
QY 241 ATTITDITQGLPEARDLLDERGNSKFTLGDMEMLDVEGVADIDPNYKFNOLLINNNALS 300
DB 241 ATTITTHITQGLPEARDLLDERGNSKFTLGDMEMLDVEGVADIDPNYKFNOLLINNNALS 300
QY 301 SVLMGSHINGIEPEKVSLLYGGNGGPGARHDWNATVGYKDQOQNNVATTINVMKNGSLV 360
DB 301 SVLMGSHINGIEPEKVSLLYGGNGGPGARHDWNATVGYKQOQNNVATTINVMKNGSLV 360
QY 361 IAGKEGINNPSFYLYKEDQITQSORALSQBEIONKIDFMEFLAQNNAKLDNLSEKEK 420
DB 361 IAGKEGINNPSFYLYKEDQITQSORALSQBEIRNKIDFMEFLAQNNAKLDNLSEKEK 420
QY 421 FRTEIKOFKDSKAYLDALGNDRITAFVSKKDTKISALITFEFGNGDLSTYLDYKKAADKA 480
DB 421 FQNEIEDFQKDSKAYLDALGNDRITAFVSKKDPKHPALITEFGKGLDSTYLLVMGKKQTEA 480
QY 481 LDREKNVTLOGSLKHGDMVEYDYSNFKYTNASKPNKGVGTNGVSHLEVGFNKAIFNL 540
DB 481 LDREKNVTLOGSLKHGDMVEYDYSNFKYTNASKPNKGVGTNGVSHLEVGFNKAIFNL 540
QY 541 PDLNLAITSFVRNRNLEKLTQGLSQPEANKLIKDFLSSNKLKELVGTLPNFNKAADAKN 600
DB 541 PDLNLAITSFVRNRNLEKLTQGLSQPEANKLIKDFLSSNKLKELVGTLPNFNKAADAKN 600
QY 601 TGNVDEVKKAQKDLKSLRKHLEKEVEKKLEKSGKNGKMKMEAKAQAQNSOKDEIFALIN 660
DB 601 TGNVDEVKKAQKDLKSLRKHLEKEVEKKLEKSGKNGKMKMEAKAQAQNSOKDEIFALIN 660
QY 661 KEANRDARAIYAONLKGIGIKRELSDKLENVKNLKDPSKDFEPKNGKNDKDFSKAEETLK 720
DB 661 KEANRDARAIYAONLKGIGIKRELSDKLENVKNLKDPSKDFEPKNGKNDKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKDFSKVTOAKSDLENSKVDIINOKV 780
DB 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKDFSKVTOAKSDLENSKVDIINOKV 780
QY 781 TDKVDNLNQAVSVAKATGDFSRVQALADLNKFSKEQLAQQAQKNSNARKKSEIYQSV 840
DB 781 TDKVDNLNQAVSVAKATGDFSRVQALADLNKFSKEQLAQQAQKNSNARKKSEIYQSV 840
QY 841 KNGVNGTLVGNGLSOAEATTLKSNFSDIKKELNAKLNKFNNNNNGLKN--EPIYAKVKN 900
DB 841 KNSVNTLVGNGLSGIEATLAKNFSDIKELNEKFNKFN--NNNNNGFNKFNSTPIYAKVKN 900
QY 901 AGQAASLEPIYAQVAKVNAKIDRLNQAISGLGVGQAGFPPL----- 944

```

```
Db 901 TGVASPEEPIYAQVAKVNAKIDRLNQAASGLGVGQ-AGFPLKRHDKVDLDSKVGRSV 959
QY 945 -----KRDHKVDLDSKVGRSVSPPIYATIDDLGGPPFLKRHDKVD 944
Db 960 SPEPIYATIDDLGGPPSKRHDKVDLDSKVGRSVSPPIYATIDDLGGPPFLKRHDKVD 1019
QY 945 -----KRDHKVDLDSKVGRSVSPPIYATIDDLGGPPFLKRHDKVD 978
Db 1020 LSKVGRSVSPPIYATIDDLGGPPSKRHDKVDLDSKVGRSVSPPIYATIDDLGGPPFLKRHDKVD 1079
QY 979 KAGFEGNLEOTIDKLKSTKHNPMLNWSAKVPASLSAKLONATNSHIRNSNLKNG 1038
Db 1080 KAGFESNLEOTIDKLKSTKHNPMLNWSAKVPASLSAKLONATNSHIRNSNLKNG 1139
QY 1039 AINEKATGMLTOKNPEWLKLVNDKIVAHNVGVSPLSEYDKIGFNQNMKDYSDSFKESTK 1098
Db 1140 AINEKATGMLTOKNPEWLKLVNDKIVAHNVGVSPLSEYDKIGFNQNMKDYSDSFKESTK 1199
QY 1099 LNNAVKDTNMGFTQFLNAFSTASYICLARENAEHGKKNVNTKGGFOKS 1147
Db 1200 LNNAVKDTNMGFTQFLNAFST-GYVSLARENAEHGKKNVNTKGGFOKS 1247

RESULT 8
ID 086064 PRELIMINARY; PRT: 1183 AA.
AC 086064;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CYTOXIN ASSOCIATED PROTEIN III.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=15818;
RA Graessle S., Redl B., Gapp G.;
RT "Molecular characterization of the cytotoxin associated gene III
RT (cagAIII) of Helicobacter pylori.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083352; AAC32880.1; -
SQ SEQUENCE 1183 AA; 132625 MW; 63968D6A651A4AB7 CRC64;

Query Match 89.7%; Score 5263; DB 2; Length 1183;
Best Local Similarity 87.3%; Pred. No. 9.1e-213;
Matches 1034; Conservative 38; Mismatches 74; Indels 38; Gaps 3;

QY 1 MTNETIDQOQTEAARNPQOQFINNLOVAFLKVDNAVASYDDQKPIVDKNDNRQAFEG 60
Db 1 MANETINQOQTEAARNPQOQFINNLOVAFLKVDNAVASYDDQKPIVDKNDNRQAFEG 60
QY 61 ISQLEEYSNKAIPKPKKQYSDFINNLOVAFLKVDNAVASYDDQKPIVDKNDNRQAFEG 120
Db 61 ISQLEEYSNKAIPKPKKQYSDFINNLOVAFLKVDNAVASYDDQKPIVDKNDNRQAFEG 120
QY 121 TSWVSHONDPSKINIRNFMENIOPPIDDKKEAEFLKSAKQSPAGIIGNOIRTDQ 180
Db 121 TSWVSHONDPSKINIRNFMENIOPPIDDKKEAEFLKSAKQSPAGIIGNOIRTDQ 180
QY 181 KFMGVDESLEKROEAKNCEPTGGDWLDFLSPFDKQSSDVKEAINOEPPVPHVOPDI 240
Db 181 KFMGVDESLEKROEAKNCEPTGGDWLDFLSPFDKQSSDVKEAINOEPPVPHVOPDI 240
QY 241 ATTTTDTIQLPPEARDLIDRGNFSEKFTLGDMEMLDVEGVADIDPNYKFNQ-LLIHNNAL 299
Db 241 ATTTTDTIQLPPEARDLIDRGNFSEKFTLGDMEMLDVEGVADIDPNYKFNQ-LLIHNNAL 300
QY 300 SSVLMGSHNGIEPKVSLLYGGNGGPGGARHDWNTATVGYKDQGGNNVATIIINVHMKNGSGL 359
```

```
Db 301 SSVLMGSHNGIEPKVSLLYAGNGGFGAKHDWNTATVGYKDQGGNNVATIIINVHMKNGSGL 360
QY 360 VIAGGERGINNPPSYLYKEDQLTQSQRALSOEEIQNKIDFMFLAQNNAKLDNLSEKEKE 419
Db 361 VIAGGERGINNPPSYLYKEDQLTQSQRALSOEEIQNKIDFMFLAQNNAKLDNLSEKEKE 420
QY 420 KFRTEIKDFQKDSKAYLDALGNDRIFAIVSKKDTKHSALITFEGNGDLSYTLKDYGGKADK 479
Db 421 KFRNEIKDFQKDSKAYLDALGNDRIFAIVSKKDTKHSALITFEGNGDLSYTLKDYGGKADK 480
QY 480 ALDREKNTVLOGSLKHGDMVVDYSNFKYTNAKPNKNGVGTNGVSHLEVGFKNKVAIFN 539
Db 481 ALDREKNTVLOGSLKHGDMVVDYSNFKYTNAKPNKNGVGTNGVSHLEVGFKNKVAIFN 540
QY 540 LPDNLNLAITSFVRRNLEDKLTTRKGLSPQEAANKLIKDFLSSNKELVGTNLNFAKAVADK 599
Db 541 LPDNLNLAITSFVRRNLEDKLTTRKGLSPQEAANKLIKDFLSSNKELVGTNLNFAKAVADK 600
QY 600 NTGNYDEVKKAQKDLKSLRREHLEKEVEKKLSKSKGNKKNMKAQAQNSOKDEIFALI 659
Db 601 NTGNYDEVKKAQKDLKSLRREHLEKEVEKKLSKSKGNKKNMKAQAQNSOKDEIFALI 660
QY 660 NKEANRDARATAYAQNLLGKIKRELSDKLENVKNLKDQKSPDEPKNGKDFSKABETL 719
Db 661 NKEASKEARAAAYVONLKGIRMEISDKLENINKDLKDFKSPDEPKNGKDFSKABETL 720
QY 720 KALGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSKVDDIINQK 779
Db 721 KALGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSKVDDIINQK 780
QY 780 VTDKVDNLNQAASVAKATGDFSRVEQALADLKNFSKEOLAQAQKNSLNARKKSEIYOS 839
Db 781 ITDKVDNLNQAASVAKATGDFSRVEQALADLKNFSKEOLAQAQKNSLNARKKSEIYOS 840
QY 840 VKNVNTLVGNGLSGIEATLAKKFSKELNEKFNKNNNNNGLNKNEPIYAKVNNK 899
Db 841 VKNVNTLVGNGLSGIEATLAKKFSKELNEKFNKNNNNNGLNKNEPIYAKVNNK 900
QY 900 KAGQAASLEERPIYAQVAKVNAKIDRLNQAASGLGVGQAGFPLKRHDKVDLDSKV--- 956
Db 901 KTGQVASPEEPIYQVAKVNAKIDRLNQAASGLGVGQAGFPLKRHDKVDLDSKVGLS 960
QY 957 -----GLSRNOELAQKIDNLNQAASVAKAGFF 983
Db 961 ANHEPIYATIDDLGGPPFLSKIRHDKVDLDSKVGLSGLNQAASVAKTCHF 1020
QY 984 GNLEQOTDKLKDSTKHNPMLNWSAKVPASLSAKLONATNSHIRNSNLKNGAINEK 1043
Db 1021 DNLDDMDLKDSTKHNPMLNWSAKVPASLSAKLONATNSHIRNSNLKNGAINEK 1080
QY 1044 ATGMLTOKNPEWLKLVNDKIVAHNVGVSPLSEYDKIGFNQNMKDYSDSFKESTKLNNAV 1103
Db 1081 ATGMLTOKNPEWLKLVNDKIVAHNVGVSPLSEYDKIGFNQNMKDYSDSFKESTKLNNAV 1140
QY 1104 KDTNMGFTQFLNAFSTASYICLARENAEHGKKNVNTKGGFOKS 1147
Db 1141 KDVKSGETHFLANAFST-GYVCLTEENAEHGKKNVNTKGGFOKS 1183

RESULT 9
ID 09F230 PRELIMINARY; PRT: 1163 AA.
AC 09F230;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
```

[illegible]

```

Db      828 LQKSVKNGVNGTLVGNSLKEATEATLTKNFSDIRKELNEKLFCGNNNNGLKNNTPEIY 887
       : : ||||| ||||| ||||| : ||||| ||||| ||||| : ||||| ||||| |||||
Qy      894 AKVNKKKACGAASLEPPIYAQVAKYNAKIDRLNQITAS-GLGVVG----- 937
       |:||||| ||| | ||||| ||||| :|||:||||| ||||| ||||| ||||| |||||
Db      888 AQVNKKKTGOATSPEEPIYAQAQAKMSAKIDRINKTASAGKYGGFSGGRSASPEIYA 947
       |:||||| ||| | ||||| ||||| :|||:||||| ||||| ||||| ||||| |||||
Qy      938 -----QAAGFLPKRHDKVDLLSKVLSRNQELAOK IDLNQAQVSEAKAGFFGNLEQTI 990
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      948 TIDFEANQAGPLRRSACGVNDSLKVLSRQEQLTRIGDLSQAVSEAKTGHDNLQEKI 1007
       |:||||| ||| | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      991 DKLKDTKHNPMLNVESAKKVPASLSAKLDNYATYNSHIRINSNIKNGATINEKATGMTQ 1050
       |:||||| ||| | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1008 DELKDSTKNALKLVESAKQVPTSLQALDNVATYNSTRINSNVQTGTINERATGMTQ 1067
       |:||||| ||| | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      1051 KPEWLKLVNDKIIVAHNVGSVPLSEYDKTGFQKNKMKDYSDSFKFTSKLNAVVKDTNSGF 1110
Db      1068 KPEWLKLVNDKIIVAHNVGSIHLSYDKITGFQKNMKDYSDSFKFTSKLNTVKDIKSSF 1127
Qy      1111 TOFLTNASTASYCYCLARENAEHGIKNVNTKGGFQKS 1147
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1128 VOFLTNTFTSGS-YSLMKAANEHGKVTNTTKGGFQKS 1163
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
Q9F229 PRELIMINARY; PRT: 1171 AA.
AC Q9F229;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CACA.
GN CACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX Helicobacter.
NCBI_TaxID=210;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=J-198;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB015407; BAB20917.1; -
SQ SEQUENCE 1171 AA; 130644 MW; E7FD06092C9E5DFE CRC64;
```

[illegible]

```
QY 298 ALSSVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNAATVGYKQDQGNVATIIIVHMKNGS 357
Db 299 ALSSVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNAATVGYKQDQGNVATIIIVHMKNGS 358
QY 358 GLVIAGGEGKINNPFSYLLKEDDQSGSALQSEETQNKIDFMEFLAQNNAKLDNLSEKE 417
Db 359 GLIIAGNEDGINKNPFSYLYKTDLGLKQALQSEETQNKIDFMEFLAQNNAKLDNLSEKE 418
QY 418 KEKFEITEIKDFKOSKAYLDALGNDRIFAVSKDKTHSALITEFNGDLSYTLKDYGGKA 477
Db 419 KEKFOFTEIENFQKDKAYLDALGNDRIFAVSKDKPKHLALVTEFNGEVSYTLKDYGGKKQ 478
QY 478 DKALDREKNVTLOGSLKHDGVNFVDSNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAI 537
Db 479 DKALDGETKTTLQGLSKYDGVNFVDSNFKYTNASKNPNKGLCATNGVSHLEAFNFKVAV 538
QY 538 FNLPLDNLNLAITSFVRNLEDKLTGKGLSPOEANKLIDFLSNKELVGKTLNFKAVAD 597
Db 539 FNLPLNLAITNYIRRDLEDKLWAGLSPOEANKLIDFLSNKELLGKVSFNFKAVAE 598
QY 598 AKNTGNYDEVKKAQKDLKSLRKRHELEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFA 657
Db 599 AKNTGNYDGVKKAQKDLKSLRKRHELEKEVEKKLESNDKNRMEAKAQAQNSQKDKIFA 658
QY 658 LINKANRDRARAIAAQNLUKIGRELSDKLENVKNLKDPSDFEFKNGKNKDFSKAAEE 717
Db 659 LINKBASKEARAAAFDPSFKIRSELSDKLENVKNLKDPSDFEFKNGKNNDFSKAAEE 718
QY 718 TLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIIN 777
Db 719 TLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSIKDVIIN 778
QY 778 QKVTDKVDNLNOAVSAKATGDFSRVEQALADLNKFSKEQLAAQQAQKNSLNARKKSEIY 837
Db 779 QKVTDKVDNLNOAVSETKLTGDFSRVEQALAEKSL-----SLDLGKNSDLQ 825
QY 838 QSVKNGVNTLVNGLSQAEATTLKSNFSDIKKELNAKLGNFNNNNNGLKNPEPIYAKVN 897
Db 826 KSVKNGVNTLVNGLSQAEATTLKSNFSDIKKELNAKLGNFNNNNNGLKNPEPIYAKVN 895
QY 898 KKKAGAAASLEPIYAOVAKV-----NAKIDRLNOIAS-GLGVVG----- 937
Db 886 KKKAGAAASLEPIYAOVAKVSAKIDQLNEATSAINRKIDRLINKIASAGKGVGFGSAG 945
QY 938 -----QAAGPPLRHRDVKVDSLKSVGLSRNQLAQKIDNLNOAVSEAKA 980
Db 946 RSASPEPIYATIDFDEANQAGPLRRAAVNDLSKVGLSREQELTRRIGDLNOAVSEAKT 1005
QY 981 GFEGNLEQTDIKDKOSTKINPNLWVESAKVPASLSAKLDNATYNSHTRINSNIKNGAI 1040
Db 1006 GHFDNLEQIDELKDKSTKKNALKLWVESAKQVPTGLQAKLDNATYNSHTRINSNVQSGTI 1065
QY 1041 NEKATGMLTKNPEWIKLVNDKIVAHNGVSPLESDYKIGFNOKNMKYSDSPKFSKTLN 1100
Db 1066 NEKATGMLTKNPEWIKLVNDKIVAHNGVSTHLYSEYDKIGFNOKNMKYSDSPKFSKTLN 1125
QY 1101 NAVKDTNSGTFQTLTNAFTASYICLARENAEHGIRKNVNTKGGFQKS 1147
Db 1126 NAVKDKISFVQTLTNTFTSGS-YSLMKANAEGHVKNNTKGGFQKS 1171
RESULT 11
Q9F218 PRELIMINARY; PRT: 1170 AA.
AC Q9F218;
DT 01-WAR-2001 (TremBLrel. 16, Created)
DT 01-WAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-WAR-2001 (TremBLrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
```

```
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J-566;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB017922; BAB20928.1; -.
SQ SEQUENCE 1170 AA; 130829 MW; 59A23543FBA86E03 CRC64;

Query Match 80.2%; Score 4703.5; DB 2; Length 1170;
Best Local Similarity 77.9%; Pred. No. 2.4e-189;
Matches 925; Conservative 99; Mismatches 106; Indels 57; Gaps 8;

QY 1 MTNETIDQ---QPQTEAANPQPFINNQLQVAFKLVDAVASYDDQKPIVDKNDNRDNRQA 57
Db 1 MTNETIDQTTTPQDT--FVPRFINNQLQVAFIKVDSAVASFDPQKPIVDKNDNRDNRQA 58
QY 58 FEGISOLREYSNKAIKNPTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFOKFGDQRY 117
Db 59 FEKISOLREYANKAIKNPAKKNOYFSDFINKSNLDINKNDLIAVDSSVESFRKFGDQRY 118
QY 118 RIFTSWVSHQNDPSKINRSIRNPMENIIQPPILDDDEKAEFLKSAKOSFAGIIIGNOIR 177
Db 119 QIFTSWVSLQKDPKSKINTQIRNFMENIIQPPISDDDEKAEFLRSAKOSFAGIIIGNOIR 178
QY 178 TDOKFMGVFDESLEKROEAEKNGEPTGGDWLDLFLSFIDKKQSSDVKEAINOEPVPHVQ 237
Db 179 SDEKFMGVFDESLEKROEAEKNAEPAGGDWLDLFLSFVFNKKQSSDLKETLQPRPDDFE 238
QY 238 PDIAATTTDTQGLPPEARLDLDERGNEFKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNN 297
Db 239 QNLATTTTDTQGLPPEARLDLDERGNEFKFTLGDVEMLDVEGVADKDPNKFENQLLIHNN 298
QY 298 ALSVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNAATVGYKQDQGNVATIIIVHMKNGS 357
Db 299 ALSSVLMGSHNSIEPEKVSLLYGGNGGPGARHDWNAATVGYKQDQGSNATLINAHLNNGS 358
QY 358 GLVIAGEKGINNPFSYLYKEDDQSGSALQSEETQNKIDFMEFLAQNNAKLDNLSEKE 417
Db 359 GLVIAGNEEGIKNPFSYLYKQDQGLKQALQSEETQNKIDFMEFLAQNNAKLDNLSEKE 418
QY 418 KEKFEITEIKDFKOSKAYLDALGNDRIFAVSKDKTHSALITEFNGDLSYTLKDYGGKA 477
Db 419 KEKFOFTEIENFQKDKAYLDALGNDRIFAVSKDKPKHLALVTEFNGEVSYTLKDYGGKKQ 478
QY 478 DKALDREKNVTLOGSLKHDGVNFVDSNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAI 537
Db 479 DKALDGETKTTLQGLSKYDGVNFVDSNFKYTNASKSPDKGVGTNGVSHLEAFNFKVAV 538
QY 538 FNLPLDNLNLAITSFVRNLEDKLTGKGLSPOEANKLIDFLSNKELVGKTLNFKAVAD 597
Db 539 FNLPLNLAITNYIRRDLEDKLWAGLSPOESNKLKIDFLSNKNKEMVGVKVSFNFKAVAE 598
QY 598 AKNTGNYDEVKKAQKDLKSLRKRHELEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFA 657
Db 599 AKNTGNYDEVKKAQKDLKSLRKRHELEKEVEKKLESNDKNRMEAKAQAQNSQKDKIFA 658
QY 658 LINKANRDRARAIAAQNLUKIGRELSDKLENVKNLKDPSDFEFKNGKNKDFSKAAEE 717
Db 659 LINKBASKEARAAAFDPSFKIRSELSDKLENVKNLKDPSDFEFKNGKNNDFSKAAEE 718
QY 718 TLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIIN 777
Db 719 TLKALKSVKDLGINPEWISKVENLNAALNDFKNGKNKDFSKVTOAKSDLENSIKDVIIN 778
QY 778 QKVTDKVDNLNOAVSAKATGDFSRVEQALADLNKFSKEQLAAQQAQKNSLNARKKSEIY 837
Db 779 QKVTDKVDNLNOAVSETKLTGDFSRVEQALAEKSL-----SLDLGKNSDLQ 825
```

```

Qy 838 QSVKGVNGTGVGLVGLSQAEATTLTKNFSDIKKELNAKLNFNNNNNNGLKNEPIYAKVN 897
Db 826 KSVKGVNGTGVGLVGLSQAEATTLTKNFSDIKKELNAKLNFNNNNNNGLKNEPIYAKVN 885
Qy 898 KKKAGQAAASLEPIYAQVAKV-----NAKIDRLNOIAS-GLGVVG----- 937
Db 886 KKKAGOVASPEPIYAQVAKVSAKIDOLNEATSAINKRIDINKIASAGKGVGFSAG 945
Qy 938 -----QAAGFPLKRHDVDDLSKVGLSRNQELAQKIDNLNOAVSEAKA 980
Db 946 RSASPEPIYATIDFDEANQAGFPLRRYAPVNDLSKVGLSRNQELTRRIGDLNQAVSEAKT 1005
Qy 981 GFFGNLEOTIDKLDKSTKHNPNLWVESAKKVPASLAKLDNYATNSHIRNSINIKNGAI 1040
Db 1006 GHFNLEOKIDELKIDSTKKNALKLWVESAKQVPTGLQAKLDNYATNSHTRNSNVQSGTI 1065
Qy 1041 NEKATGMLTQKNPEWLKLVNDKIVAHNVGVSPLSEYDKIGFNOKNMKDYSDSEKFFSTKLN 1100
Db 1066 NEKATGMLTQKNPEWLKLVNDKIVAHNVGVSPLSEYDKIGFNOKNMKDYSDSEKFFSTKLN 1125
Qy 1101 NAVKDTNSGFTQFTLNAFSTASYCYCLARENAEHGKINVTNRKGGFQKS 1147
Db 1126 NAVKDIKSSFFQVLTNTFSTGS-YSLMKANVEHGVEN-TTRSGFQKS 1170

RESULT 12
Q9RF16 PRELIMINARY; PRT; 1173 AA.
AC Q9RF16
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=GC401(F32);
RA Ito Y., Azuma T.
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF202972; AAF17597.1; -
SQ SEQUENCE 1173 AA; 130940 MW; C36709DCF2946209 CRC64;

Query Match 80.1%; Score 4700; DB 2; Length 1173;
Best Local Similarity 77.6%; Pred. No. 3.3e-189;
Matches 921; Conservative 98; Mismatches 114; Indels 54; Gaps 7;

Qy 1 MTNETIDQ-QPQTEAFNPOQFINNLOVAFKVDNNAVASYDPQOKPIVDKNDNRNQAFE 59
Db 1 MTNETIDQTPQGTGVPQRFINNLOVAFKVDNNAVASYDPQOKPIVDKNDNRQAYE 60
Qy 60 GISOLREYNKAIKNTKKNQVPSDFINKSNDLINDVESSTKSFQKFGDQRYI 119
Db 61 KISOLREYANKAIKNTKKNQVPSDFINKSNDLINDVESSTKSFQKFGDQRYI 120
Qy 120 FTSWVSHQNDPSKINTRSINRFENIITQPTLDDKEAEFLSAKOSFAGIIGNQIRTD 179
Db 121 FTSWVSLQKQPSKINTQIIRNFENIITQPTLDDKEAEFLSAKOSFAGIIGNQIRSD 180
Qy 180 QKFMGVFDESILKQAEKNGEPTGGDWLDIFLSFIFDKQSSDVKEAINQEPVPHVQPD 239
Db 181 EKFMGVFDESILKQAEKNAEPAGGWDLDIFLSFVFNKQSSDLKETLNQEPDPDEQN 240
Qy 240 IATTTDIOGLPPEARDLLBERGNSFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNAL 299
Db 241 LATTTDIOGLPPEARDLLBERGNSFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNAL 300
Qy 300 SSVLMGSHNGIEPEKVSLLYGGNGGPGARHDWATVGYKQOQGNVATILNVHMKNGSGL 359

```

```

Db 301 SSMLMGSHNSHIEPEKVSLLYGGNGGPGARHDWATVGYKQOQGNVATILNAHLNNGSGL 360
Qy 360 VIAGKEGGINNPSFYLYKREDQLTGQSORALSQOEIQNKIDFMEFLAONNAKLNDLSEKEKE 419
Db 361 IAGNEDGINKPSFYLYKREDQLTGLQALSQOEIQNKIDFMEFLAONNAKLNDLSEKEKE 420
Qy 420 KPRTEIKDFQKDSKAYLDALGNDRIFAFVSKKDTKHSALITEFGNGDLSYTLKDYGKKADK 479
Db 421 KFOTEIENFQDKRAYLDALGNDRIFAFVSKKDKPHLALVTEFGNGELSYYTLKDYGKKQDK 480
Qy 480 ALDRKKNVTLOGSLKHDGVMFVDYSNFKYTNASKNPNKGVGTNGVSHLEVCFNKVAITN 539
Db 481 ALDGETKTTLOGSLKYDGMFVFNYSNFKYTNASKNPNKGVGTNGVSHLEVCFNKVAITN 540
Qy 540 LPDLNLAIITFVRNRNLEDKLTGKLSPOEANKLLKDFLSSNKELVGKTLNFKNAVADAK 599
Db 541 LPNLNLAIITVIRDLDEKLLWAKGLSPOEANKLLKDFLSSNKELVGKTLNFKNAVADAK 600
Qy 600 NTGNYDEVYKKAQKDLKSLRKRREHLEKEVEKLESKSGNKNKMEAKAQAANSOKEIFALI 659
Db 601 NTGNYDEVYKKAQKDLKSLRKRREHLEKEVEKLESKSGNKNKMEAKAQAANSOKDKIFALI 660
Qy 660 NKEANRDARATAYAQNLKIGIKRELSDKLENVNKNLKDQKSFDEPFKNKNKDFSKAEETL 719
Db 661 SQAASKEARVATFDPYLGKVRSELSDKLENVNKNLKDQKSFDEPFKNKNKDFSKAEETL 720
Qy 720 KALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKSKYTOAKSDLENSVKDVLINOK 779
Db 721 KALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKSKYTOAKSDLENSVKDVLINOK 780
Qy 780 VTDKVDNLNQAVSAKATGDFSRVQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQS 839
Db 781 ITDKVDNLNQAVSEIKLTGDFSKVQALAEKLN-----SIDLGKNSDLQKS 827
Qy 840 VKNGVNTLVNGLSQAEATTLTKNFSDIKKELNAKLNFNNNNNNGLKNEPIYAKVN 897
Db 828 VKNGVNTLVNGLSQAEATTLTKNFSDIKKELNAKLNFNNNNNNGLKNEPIYAKVN 887
Qy 898 KKKAGQAAASLEPIYAQVAKV-----NAKIDRLNOIAS-GLGVVG----- 937
Db 886 KKKAGOVASPEPIYAQVAKVSAKIDOLNEATSAINKRIDINKIASAGKGVGFSAG 947
Qy 938 -----QAAGFPLKRHDVDDLSKVGLSRNQELAQKIDNLNOAVSEAKA 980
Db 946 RSASPEPIYATIDFDEANQAGFPLRRYAPVNDLSKVGLSRNQELTRRIGDLNQAVSEAKT 1007
Qy 981 GFFGNLEOTIDKLDKSTKHNPNLWVESAKKVPASLAKLDNYATNSHIRNSINIKNGAI 1040
Db 1006 GHFNLEOKIDELKIDSTKKNALKLWVESAKQVPTGLQAKLDNYATNSHTRNSNVQSGTI 1067
Qy 1041 NEKATGMLTQKNPEWLKLVNDKIVAHNVGVSPLSEYDKIGFNOKNMKDYSDSEKFFSTKLN 1100
Db 1066 NEKATGMLTQKNPEWLKLVNDKIVAHNVGVSPLSEYDKIGFNOKNMKDYSDSEKFFSTKLN 1127
Qy 1101 NAVKDTNSGFTQFTLNAFSTASYCYCLARENAEHGKINVTNRKGGFQKS 1147
Db 1126 NAVKDIKSSFFQVLTNTFSTGS-YSLMKANVEHGVEN-TTRSGFQKS 1173

RESULT 13
Q9F231 PRELIMINARY; PRT; 1176 AA.
AC Q9F231
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=J-149;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RT Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
Japanese and non-Japanese isolates."
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL: AB015405; BAB20915.1;
SQ SEQUENCE 1176 AA; 131209 MW; DA8D473D2E2C47EE CRC64;

Query Match 79.9%; Score 4690.5; DB 2; Length 1176;
Best Local Similarity 77.6%; Pred. No. 8.4e-189;
Matches 928; Conservative 87; Mismatches 117; Indels 59; Gaps 8;

QY 1 MTNETIDQ-----QPOTEAFAFNPOQFINNLOVAFKVDNAVSDPDQKPIVDKNDNRN 55
DB 1 MTNETIDQTTPTDTPNQTDFVQRFINNLOVAFKVDNAVSDPDQKPIVDKNDNRN 60

QY 56 QAPEGISQLEEEYANKAIPKTKKNOYFSDFINKSNLDINKONLIDVESKTSFQKFGDQ 115
DB 61 QAPEKISQLEEEYANKAIPKTKKNOYFSDFINKSNLDINKONLIDVESKTSFQKFGDQ 120

QY 116 RYRFTSWSHQNDPSKINTSRNFMENIIOPIIDDDKEKAEFLSAKOSFAGIIGNQ 175
DB 121 RYQIFTSWLSQKDPKSKINTQIRNFMENIIOPIIDDDKEKAEFLSAKOSFAGIIGNQ 180

QY 176 IRTDOKFMGVFDESLEKROEAEKNGEPTGGDWLIDFLSFIEDKKOSSDVKEAINEQPVPH 235
DB 181 IRSDEKFMGVFDESLEKROEAEKNGEPTGGDWLIDFLSFIEDKKOSSDVKEAINEQPVPH 240

QY 236 VQPIATTTTIDQGLPPPEARLDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIIH 295
DB 241 FEQNLATTTTIDQGLPPPEARLDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIIH 300

QY 296 NNALSSVLMGSHNGIEPEKVSLLYGGNGPGARHDNATVGYKDOQGNVATIIINVHKN 355
DB 301 NNALSSVLMGSHNGIEPEKVSLLYGGNGPGARHDNATVGYKDOQGNVATIIINVHKN 360

QY 356 GSGVLVAGEGKGNFNFYLYKEDQGTGSRQALSDQEIQKIDFMEFLAONNAKLNLSE 415
DB 361 GSGVLVAGENGKGNFNFYLYKEDQGTGSRQALSDQEIQKIDFMEFLAONNAKLNLSE 420

QY 416 KEKEKETEIKDFQKDSKAYLDALGNDRIFAIVSKKDTKHSALITEFGNGDLSYTLKDYGK 475
DB 421 KEIRKPESEIENFQKNPKAYLDALGNDRIFAIVSKKDTKHSALITEFGNGDLSYTLKDYGK 480

QY 476 KADKALDREKNVTLOGSLKHGDMVFDVSNFKYTNASKNPKNGVGTNGVSHLEVGKVV 535
DB 481 KQKALDGETKTTLOGSLKYDGMVFNYSNFKYTNASKNPKNGVGTNGVSHLEAFNFSKV 540

QY 536 AIFNLPLNLALITSFVRRLNLEKLTTKGLSPQEAANKLIKDFLSSNKKELVGTLLNFKAV 595
DB 541 AVENLPLNLALITNIRDLDEKLUAKGLSPQEAANKLIKDFLSSNKKELVGTLLNFKAV 600

QY 596 ADAKNTGNYDEVKKAQKLESLRKEHLEKEVEKKLESKSGNKNMEAKAQAQNSQKDEI 655
DB 601 AEAKNTGNYDEVKKAQKLESLRKEHLEKEVEKKLESKSGNKNMEAKAQAQNSQKDEI 660

QY 656 FALINKEANDRAIAYAQNKGKRELSDKLENVKNLKDFOKSFDEFKNGKNKDESKA 715
DB 661 FALINTEASKEARVAFDPLNKGIRSELSDKLENINKNLKDFOKSFDEFKNGKNKDESKA 720

QY 716 EETLKALKGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSKVDVI 775
DB 721 EETLKALKGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSKVDVI 780

QY 776 INOKVTDKVDNLNQAVSVKATGDFSRVEQALADLNFSKEQLAQAQKKNESLNARKKSE 835
DB 781 INQKITDKVDNLNQAVSETKLTGDFSKVEQALAEKLSL-----SLDLGKNSD 827

QY 836 IYQSVKNGVNTLVGNSLQSAEATTTLSKNFSDIKKELNAKLGNFNNNNNGLKN--EPIY 893
DB 836 IYQSVKNGVNTLVGNSLQSAEATTTLSKNFSDIKKELNAKLGNFNNNNNGLKN--EPIY 893

DB 828 LQKSVKNGVNTLVGNSLQSAEATTTLSKNFSDIKKELNAKLGNFNNNNNGLKNSTPIY 887

QY 894 AKVKKKAGQAASLEEEPIYAQVAKV-----NAKIDRLNQIAS-GLGVVG- 937
DB 888 AKVKKKAGQAASLEEEPIYAQVAKV-----NAKIDRLNQIAS-GLGVVG- 947

QY 938 -----QAAGFPLKHKDKVDDLSKVLGSRNQELAQKIDNLNOAVS 976
DB 948 SGAGRSASPEPIYATIDFDEANQAGFLRRSTGVNDLSKVLGSRNQELTRIGDLNOAVS 1007

QY 977 EAKAGFGNLEOTIDKLDSTKHNPMNLWVESAKKVPASLASKLDNATNSHIRINSNIK 1036
DB 1008 EAKTGYDNLEQKIDELKSTKKNALKLVESAKQVPTGLOAKLDNATNSHIRINSNH 1067

QY 1037 NGAINKATGMLTKQKPEWLKLVNDKIVAHNVGSPVLESEYDKIGFNQKMKDYSDSEKFS 1096
DB 1068 NGAINKATGMLTKQKPEWLKLVNDKIVAHNVGSAHLSEYDKIGFNQKMKDYSDSEKFS 1127

QY 1097 TKLNAVKTNSGFTQPLTNAFSTASYCLARENAEHGKIKNVNTKGGFQKS 1147
DB 1128 TKLNAVKTNSGFTQPLTNAFSTASYCLARENAEHGKIKNVNTKGGFQKS 1176

RESULT 14
Q9F228 PRELIMINARY; PRT; 1176 AA.

AC Q9F228;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE CAGA.
GN CAGA.

OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
ON NCBI_TaxID=210;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J-207;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
Japanese and non-Japanese isolates."
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL: AB015408; BAB20918.1;
SQ SEQUENCE 1176 AA; 131015 MW; 9842E72146F61B1B CRC64;

Query Match 79.9%; Score 4690.5; DB 2; Length 1176;
Best Local Similarity 77.6%; Pred. No. 8.4e-189;
Matches 924; Conservative 93; Mismatches 115; Indels 59; Gaps 8;

QY 1 MTNETIDQ-----QPOTEAFAFNPOQFINNLOVAFKVDNAVSDPDQKPIVDKNDNRN 55
DB 1 MTNETIDQTTPTDTPNQTDFVQRFINNLOVAFKVDNAVSDPDQKPIVDKNDNRN 60

QY 56 QAPEGISQLEEEYANKAIPKTKKNOYFSDFINKSNLDINKONLIDVESKTSFQKFGDQ 115
DB 61 QAPEKISQLEEEYANKAIPKTKKNOYFSDFINKSNLDINKONLIDVESKTSFQKFGDQ 120

QY 116 RYRFTSWSHQNDPSKINTSRNFMENIIOPIIDDDKEKAEFLSAKOSFAGIIGNQ 175
DB 121 RYQIFTSWLSQKDPKSKINTQIRNFMENIIOPIIDDDKEKAEFLSAKOSFAGIIGNQ 180

QY 176 IRTDOKFMGVFDESLEKROEAEKNGEPTGGDWLIDFLSFIEDKKOSSDVKEAINEQPVPH 235
DB 181 IRSDEKFMGVFDESLEKROEAEKNGEPTGGDWLIDFLSFIEDKKOSSDVKEAINEQPVPH 240

QY 236 VQPIATTTTIDQGLPPPEARLDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIIH 295
DB 241 FEQNLATTTTIDQGLPPPEARLDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIIH 300

QY 296 NNALSSVLMGSHNGIEPEKVSLLYGGNGPGARHDNATVGYKDOQGNVATIIINVHKN 355


```
Db 301 NNALSSVLMGSHSIEPEKVSLSLYGONGGPEARHDMNATVGYKQQGNNVATLINVHLNN 360
Qy 356 GSGLVTAGGEGKGNPNPFIYKEDQTLGQSORALSQOEIQNKIDFMEFLAQNNAKLDNLSE 415
Db 361 GSGLIAGNEDGKGNPFIYKEDQTLGKQASQOEIQNKIDFMEFLAQNNAKLDNLSE 420
Qy 416 KEKEKFEITEKDFOKSKAYLDALGNDRIAFVSKDKTHSALITEFGNGDLSYTLKDYGK 475
Db 421 KEKEKFEITEKDFOKSKAYLDALGNDRIAFVSKDKTHSALITEFGNGDLSYTLKDYGK 480
Qy 476 KADKALDREKNVTLOGSLKHGDMVFNYSNFKYTNASKNPNKNGVGTNGVSHLEVGNKV 535
Db 481 KQKALDGEIKTTLQSLKSLYKDGVMFNYSNFKYTNASKNPNKNGVGTNGVSHLEVGNKV 540
Qy 536 AIFNLPDLNLATISFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGTNLNFKAV 595
Db 541 AVFNLPDLNLATISFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGTNLNFKAV 600
Qy 596 ADAKNTGNYDEVKKAQKDLKSLRKRHEHLEKEVEKLESKSGNKNMEAKQAQNSQKDEI 655
Db 601 AEAKNTGNYDEVKKAQKDLKSLRKRHEHLEKEVEKLESKSGNKNMEAKQAQNSQKDEI 660
Qy 656 FALINKEANDARAIAQAQNLKGIKRELSDKLENNVKNLKDPSKDEFEKNGKNKDFSKA 715
Db 661 FALINQEAKEARAAAFDPSPKIRSELSDKLENINKLKDGLSKDELKNGKNDFSKA 720
Qy 716 EETLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTKQAQSDLENSYKDV 775
Db 721 EETLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTKQAQSDLENSYKDV 780
Qy 776 INQKVTDKVDNLNOQAVSAKATGDFSVEQALADLKNFSKEQLAQQAQKNESLNARKKSE 835
Db 781 INQKITDKVDNLNOQAVSETKLADGFSKEQLAELKSL-----SLDLGKNSD 827
Qy 836 IYOSVKNVNGTLVGNGLSQAEATTLKSNFSDIKKELNAKLGNNNNNNGLKN--EPIY 893
Db 828 LQSVKNGVNGTLVGNGLSKTEATTLKSNFSDIKKELNAKLGNNNNNNGLKNSAEPIY 887
Qy 894 AKYNKKKAGQAASLEPIYAQVAKV-----NAKIDRLNQIAS--GLGVVG- 937
Db 888 AQVKKKTKQVASEPIYAQVAKVSAKIDQLNEATSAINKRIDRLNK TASAGKGVGGF 947
Qy 938 -----QAAGPLKRHDVDDLKSVGLSRNOELAQKIDNLNOQAVS 976
Db 948 SGAGRSASPEPIYATIDFDEANQAGFPLRRSAGVNDLSKVGLSREOELTRRIGLSQAVS 1007
Qy 977 EAKAGFGLNEQTLKLDSTKINPNLWVESAKKVPASLSAKLDNYATNSHIRINSNIK 1036
Db 1008 EAKTGHFDNLEQKIDELKSTKKNALKLWVESAKQVPTGLQAKLDNYATNSHIRINSNVQ 1067
Qy 1037 NGAINKEKATGMLTKQNPWELKLVNDKIVAHNVGVSPLSEYDKIGFNKNMKDYSDGPKFS 1096
Db 1068 TGFINKEKATGMLTKQNPWELKLVNDKIVAHNVGVSAPLSEYDKIGFNKNMKDYSDGPKFS 1127
Qy 1097 TKLNNVADPNSTGFTOFLTNAFSTASYICLARENAEHGKINVTNKGQFQKS 1147
Db 1128 TKLNNVADIKSSFVQFLTNTSTGFS-YSLMKANAHEGVKN--TTKSGFQKS 1176

RESULT 15
Q9F225 PRELIMINARY; PRT; 1177 AA.
AC Q9F225;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
```

```
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=J-241;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RL J. Gastroenterol. 35:890-897(2000).";
DR EMBL: AB015411; BAB20921.1; -.
SQ SEQUENCE 1177 AA; 131346 MW; EIBC418998F410B5 CRC64;

Query Match 79.9%; Score 4689; DB 2; Length 1177;
Best Local Similarity 77.6%; Pred. No. 9,7e-189;
Matches 924; Conservative 94; Mismatches 115; Indels 58; Gaps 7;

Qy 1 MTNETIDQ-----QPQTEAAFNPOQFINNLOVAFKLVYDNAVASYDPQKPIVDKNDNR 55
Db 1 MTNETIDQTTTPDQTPNQDTFVPOQFINNLOVAFIKVDSAVASFPDQKPIVDKNDNR 60
Qy 56 QAPEGISOLREESYKAIKNPTKKNOYFSDPINKSNDLINKNDLINKNDLVESSTKSFQKFGDQ 115
Db 61 QAPEKISOLREEVANKAIKNPAKANOYFSDPINKSNDLINKNDLINKNDLVSSVESRKFQDQ 120
Qy 116 RYRIFTSWVSHQNDPSKINTSRIRNFMENIIPQILDDEKAEFLKSAKQSFAGIIGNQ 175
Db 121 RYLNFTSWVSLQKDPKINTQIQONFENIIPQISDDKEAEFLRSKQSFAGIIGNQ 180
Qy 176 IRDOKFMGVFDESLEKROAEKNGEPTGGDWLDFISFIDFKQSSDVKCAINQEPVPH 235
Db 181 IRSDEKFMGVFDESLEKROAEKNEQAGGDMWDFISFVFNKQSSDLKETLQOEPRPS 240
Qy 236 VQPDIAITTTDIOGLPPEARDLIDERNFSGFTLGDMEMLDVEGVADIDPNKFNOLLIH 295
Db 241 VEQNIATTTDIOGLPPEARDLIDERNFSGFTLGDVEMLDVEGVDPKDPNKNOLLIH 300
Qy 296 NNALSSVLMGSHNGIEPEKVSLSLYGNGGPGARHDMNATVGYKQQGNNVATLINVHLNN 355
Db 301 NNALSSVLMGSHSIEPEKVSLSLYGNGGPGARHDMNATVGYKQQGNNVATLINVHLNN 360
Qy 356 GSGLVTAGGEGKGNPNPFIYKEDQTLGQSORALSQOEIQNKIDFMEFLAQNNAKLDNLSE 415
Db 361 GSGLIAGNEDGKGNPFIYKEDQTLGKQASQOEIQNKIDFMEFLAQNNAKLDNLSE 420
Qy 416 KEKEKFEITEKDFOKSKAYLDALGNDRIAFVSKDKTHSALITEFGNGDLSYTLKDYGK 475
Db 421 KEKEKFEITEKDFOKSKAYLDALGNDRIAFVSKDKTHSALITEFGNGDLSYTLKDYGK 480
Qy 476 KADKALDREKNVTLOGSLKHGDMVFNYSNFKYTNASKNPNKNGVGTNGVSHLEVGNKV 535
Db 481 KQKALDGEIKTTLQSLKSLYKDGVMFNYSNFKYTNASKNPNKNGVGTNGVSHLEVGNKV 540
Qy 536 AIFNLPDLNLATISFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGTNLNFKAV 595
Db 541 AVFNLPDLNLATISFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGTNLNFKAV 600
Qy 596 ADAKNTGNYDEVKKAQKDLKSLRKRHEHLEKEVEKLESKSGNKNMEAKQAQNSQKDEI 655
Db 601 AEAKNTGNYDEVKKAQKDLKSLRKRHEHLEKEVEKLESKSGNKNMEAKQAQNSQKDEI 660
Qy 656 FALINKEANDARAIAQAQNLKGIKRELSDKLENNVKNLKDPSKDEFEKNGKNKDFSKA 715
Db 661 FALINQEAKEARAAAFDPNPKIRSELSDKLENINKLKDGLSKDELKNGKNDFSKA 720
Qy 716 EETLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTKQAQSDLENSYKDV 775
Db 721 EETLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTKQAQSDLENSYKDV 780
Qy 776 INQKVTDKVDNLNOQAVSAKATGDFSVEQALADLKNFSKEQLAQQAQKNESLNARKKSE 835
Db 781 INQKITDKVDNLNOQAVSETKLADGFSKEQLAELKSL-----SLDLGKNSD 827
Qy 836 IYOSVKNVNGTLVGNGLSQAEATTLKSNFSDIKKELNAKLGNNNNNNGLKN--EPIY 893
```


Search completed: February 13, 2002, 14:11:58
Job time: 657 sec

THIS PAGE BLANK (USPTO)